

Statement of Dr. Frederick W. Lipfert
Department of Applied Science
Brookhaven National Laboratory
Upton, NY 11973

before the U.S. House of Representatives Commerce Committee on
May 8, 1997

Good morning and thank you for the opportunity to appear at this hearing. First, a few words of introduction. I am appearing here today as a member of the scientific staff of Brookhaven National Laboratory, located on Long Island at Upton, New York, where I have worked for the past 18 years, in both research and management positions. Earlier in my career, I did research in the aerospace industry and served as manager of air quality for an electric utility company. I hold professional degrees in engineering and fluid mechanics and a Ph.D. in environmental studies. My book on air pollution epidemiology was published in 1994.

I should also tell you that, for the past 12 years, I have worked at Brookhaven part-time in order to establish a professional consulting practice. My clients have included both government and industry, as well as environmentalists. For example, I was retained by the U.S. EPA to prepare material for the recent Criteria Document on particulate matter (Section 12.4 on long-term mortality effects). At present, I am working for the Electric Power Research Institute on various air quality issues and for the Canadian government on an analysis of the potential benefits of reducing sulfur in motor vehicle fuels.

Recently, the U.S. Department of Energy requested Brookhaven to provide comments on EPA's proposed revisions to the National Ambient Air Quality Standards for particulate matter, which is one of the subjects of this hearing. The report that I then prepared for DOE comprises the technical portion of my testimony.

Because of the highly technical nature of that report, I would like to summarize some of its findings for you. Since I only have 5 minutes, these points must be brief, but I would be happy to elaborate later on any topics of particular interest.

- o The proposed EPA standards are based entirely upon epidemiological (i.e., statistical) studies. There has been no experimental confirmation of health effects at these low concentration levels, nor have plausible biological mechanisms been identified.
- o The findings of these studies depend on the statistical methods used. In each case in which a study has been reanalyzed by a different investigator, different results have been found.
- o EPA has selected only a portion of the epidemiological record for emphasis (seven studies). Newer studies do not support their position.

- o These studies cannot reliably distinguish which air pollutant or type of particle may actually be responsible. This results in part from the fact that most pollutants are correlated with one another, that some of them (fine particles) tend to be measured more reliably than others, and that all pollutant candidates have not been considered in EPA's studies.
- o Confounding variables have not been properly considered. Problem areas include weather effects in the short-term studies and regional lifestyles differences in the long-term studies.
- o None of these statistical studies can identify the individuals that may have been affected by air pollution or determine what their actual exposures may have been. This lack of knowledge precludes the determination of threshold concentrations that might support the setting of rational standards.

For these reasons, I believe that the scientific and technical support for EPA's proposed new standards is inadequate and that the proposal should be reconsidered or postponed.

Finally, I would like to take advantage of this opportunity to speak to you directly and personally about the current air pollution control situation in the U.S. We now enjoy perhaps the best air quality of any major industrial nation, but, by and large, the easy solutions have all been implemented. Further improvements will be expensive and perhaps intrusive, and these burdens will be borne by all of us. Thus, the current debate is appropriate and useful, and I am grateful that the Congress is facilitating at least a portion of it. One aspect that has not received much attention is the need to look back with a retrospective epidemiological study to determine just what health benefits we have actually gained as a result of the dramatic improvements in air quality over the past 30 years (which include major reductions in fine particles, by the way). Essentially, we have conducted an intervention trial on 200 million people, but no one has yet taken the time to evaluate the results. How can we steer the right course for the future if we don't know where we've been?

As an independent scientist, my plea to the Congress is for the wherewithal to bring the research that has been started to a meaningful resolution; in other words, give us the tools to finish the job. One of the critical needs at the moment is access to data, from government agencies and from research entities that were supported with public funds. As EPA expands its ambient air quality monitoring base (which should have had a higher priority than promulgating new regulations), this situation should improve, but it will probably take 5 years to accomplish. And perhaps most importantly, the Federal government should step up to the plate and take charge of the research agenda with a multi-agency approach, perhaps similar to what was done with acid rain in the 1980s. If we are truly facing a significant public

health problem, then health-care agencies should be directly involved, as well as those agencies concerned with implementing pollution controls. In this large sense, establishing the proper national direction is in your hands. I sincerely hope we can all work together for the good of the nation.

Thank you for listening.

EPA based its proposal for a new air quality standard for fine particles on its judgment that public health would not be adequately protected by the existing air quality standards that are based on a more comprehensive index of airborne particles. EPA cited seven epidemiological studies in support of this position, three based on daily (acute) responses, and four based on longer time periods (a year or more). These studies are listed on p. 48 of my report. None of these studies was able to identify a "safe" level of air quality, i.e., a threshold concentration above which health effects might be expected. EPA's proposed standards were thus based on judgment and have been set at a level that could result in severe restrictions on economic growth in the most highly populated areas of the nation. They also have the potential to result in regional factionalism, i.e., competition between the areas characterized by fine particles and those areas in which coarser airborne particles are more typical.

The draconian nature of EPA's proposal has thus set both political and scientific debates into motion. My report was confined to scientific issues, which I will outline for you, with some additional elaboration.

First, on the short-term or daily epidemiological studies. These and all of the relevant epidemiological studies are seriously limited by the availability of appropriate air quality data. It should be obvious that responses to air pollution cannot be evaluated properly without knowing what levels people are actually exposed to. Obtaining this information is not a simple matter, especially since all of the health responses linked to air pollution have many other (and usually, much more important) causes besides air pollution. Asthma is a case in point; a recent guideline from the New England Journal of Medicine (which cited the American Lung Association as a source) advised asthmatics to avoid "molds, household dust mites, and second-hand tobacco smoke." Outdoor air pollution was not mentioned, which is consistent with the fact that most people spend more than 90% of their time indoors and thus that indoor air pollution sources must also be considered when evaluating health responses.

The situation becomes more critical when one tries to differentiate health responses according to particle size, a distinction that lies at the heart of the EPA proposal. First, the sampling devices on which these distinctions are based tend to favor fine

over coarse particles. The equipment used in at least two of the studies used 90% of the sampled air for fine particles and only 10% for coarse particles; this resulted in much more accurate and precise data for the fines than for the coarse, by about a factor of 3. To make matters even worse, the key EPA short-term study relied upon a single air sampler in each of six metropolitan areas as large as 9 counties. Since the smallest particles remain airborne much longer than coarse particles, fines are distributed much more uniformly across a city than coarse particles, leading to a further accuracy bias in favor of the fines. Furthermore, in several of the areas, the air sampler was not located in the densest part of the metropolitan area, so that the actual population exposures to coarse particles may have been badly underestimated. Now, it is well known to statisticians

that poorly measured exposures lead to biased estimates of their effects, usually biased downward. Thus, the key EPA short-term study very likely resulted in a "false-negative" for coarse particles. We should also note that the most polluted area that was study showed statistical significance and a larger effect for coarse particles than it did for fine, in spite of the substantial measurement error handicap that I just outlined. This information appears in much more detail in a paper that was published in last month's issue of the Journal of the Air and Waste Management Association.

In the study of daily hospital admissions for respiratory causes that EPA relied upon, the sum of fine and coarse particles showed a larger mean effect than did fine particles alone.

Now, we turn to longer-term studies, which are designed to evaluate whether exposure to air pollution results in lasting effects on health, as opposed to short-term exacerbations. By all accounts, this is a much more difficult task, involving some of the same air quality measurement problems but also a set of different problems. First, it is important to distinguish a response that involves creating a new case of chronic disease, such as chronic bronchitis or coronary heart disease, from one that exacerbates existing symptoms on time-scales longer than within a few days. In order to make this determination, exposures to air pollution must be considered during the period before symptoms of the disease actually appear, because of the length of time typically required for development. Only one study has gone to this level of detail, involving a group of Seventh-Day Adventists in California. Exposures of these people to a wide variety of air pollutants were determined on an individual basis for 10 years before follow-up began (it is my understanding that this research is continuing), and the best estimates for mortality from all causes and for new cardiac problems were that of no effect. The study did find that previous exposures to particles of all sizes at levels in excess of the present standard resulted in some new cases of respiratory disease.

However, this is not one of the studies that EPA has chosen to rely upon. Instead, their two selected mortality studies used air quality data essentially contemporary with the period of follow-up; that is, deaths occurring during this period were

linked to the average air quality of the period, and not to previous exposures. Furthermore, individual exposures were not determined or estimated; all residents of each community were assumed to have the same exposure, resulting in the same particle-size bias seen in the daily studies.

In my book in 1994, I reviewed the epidemiology that had then been published and "called for action" on both scientific and political fronts. I can't take credit for the tremendous spurt in the science that has since ensued, thus I feel free to criticize its lack of direction and focus. Much of it has been repetitive, some of it has been self-serving, and it is perhaps

symptomatic of the current status that each reanalysis of an existing study has produced results that differ from the original. One is led to the conclusion that there are indeed multiple answers to these questions and that no one has a monopoly on the "truth." Without an expanded and more focused research program, we are very likely to find ourselves back here again within a few years, either wrestling with a different air pollutant or else responding to a legal challenge that the standards should be tightened still further.

; I was heavily involved in the National Acid Precipitation Assessment Program, for example.

Technical Comments on EPA's Proposed Revisions to the National Ambient Air Quality Standard for Particulate Matter

prepared for the U.S. Department of Energy, Office of Fossil
Energy

Executive Summary

Background. The U.S. Environmental Protection Agency (EPA) has proposed new ambient air quality standards specifically for fine particulate matter, regulating concentrations of particles with median aerodynamic diameters less than 2.5 μm (PM_{2.5}). Two new standards have been proposed: a maximum 24-hr concentration (50 $\mu\text{g}/\text{m}^3$) that is intended to protect against acute health effects, and an annual concentration limit (15 $\mu\text{g}/\text{m}^3$) that is intended to protect against longer-term (chronic) health effects. EPA has also proposed a slight relaxation of the 24-hr standard for inhalable particles (PM₁₀), by allowing additional exceedances each year (150 $\mu\text{g}/\text{m}^3$). These proposals were published in the Federal Register (FR 61:65638, Dec. 13, 1996) and are the result of a lawsuit that mandated EPA action under a court-defined schedule.

Fine particles are currently being controlled indirectly by means of regulations for PM₁₀ and TSP, under the Clean Air Act of 1970 and subsequent amendments. Although routine monitoring of PM_{2.5} is rare and data are sparse, the available data indicate that ambient concentrations have been declining at about 6% per year under existing regulations.

The proposed NAAQS revisions are not entirely consistent with the recommendations of EPA's Clean Air Scientific Advisory Committee (CASAC). Only two of the 21 CASAC members recommended an annual PM_{2.5} standard as low as the 15 $\mu\text{g}/\text{m}^3$ level that EPA proposed and that eight of them felt that no such annual standard was appropriate. In addition, the proposed NAAQS revisions may be unique in the history of the Clean Air Act, in that they are based almost solely on epidemiology, none of the epidemiological studies is capable of identifying a "safe" exposure level, and no plausible biological mechanisms have been identified that would support the epidemiological findings.

Technical Issues Pertaining to Epidemiological Studies of Both Short- and long-Term Health Effects

1. Health effects specific to fine particles. At the heart of the EPA proposals for new NAAQS for PM_{2.5} is their contention that health responses have been determined that are specific to this class of particles, from which the public would not be protected by means of controls on PM₁₀. EPA cites seven epidemiological studies in support of this position, three for acute effects and four for long-term effects. First, there is no health end point that is unique to fine particles, to any type of particle, or to air pollution in general; all of the health effects are multifactoral, responding to a variety of environmental and non-environmental stimuli. Second, based on

the studies that EPA has selected, in no case does the mean health response associated with PM_{2.5} significantly exceed that associated with PM₁₀; in most cases, PM₁₀ effects are greater and the effects attributed to sulfate (SO₄²⁻ is an important component of both PM₁₀ and PM_{2.5}) are a distant third. As a result, there is no health effects basis for a separate PM₁₀ standard.

2. Pollutant collinearity. Mixtures of air pollutants are found in all urban areas, because of common sources. In addition, many pollutants respond in concert to changes in weather. In epidemiological analysis, failure to consider all relevant exposures to air pollution and to weather extremes will result in overstating the responses to those agents that are considered. For example, during hot weather, elevated PM, ozone (O₃), and temperature levels tend to occur together. Many of the EPA-selected studies considered only a few pollutants; carbon monoxide (CO) has been a common omission, for example. To the extent that CO and particulate matter (PM) are correlated (and the PM from vehicles tends to be highly correlated with CO), the health effects of PM are likely to have been exaggerated or mis-stated. Recent studies of hospital admissions have shown that O₃ is more important than PM for respiratory diseases and that CO is more important than PM for certain cardiac diagnoses. Another mortality study showed that inclusion of relative humidity in the model eliminated PM as a significant risk factor.

3. Uncertainties in the air pollution levels that the putative victims of air pollution were actually exposed to. All epidemiological studies use air quality data as obtained from one or more centrally located air monitoring stations in each community, as opposed to direct measurements of individual exposures. The limited studies of personal exposures show poor correlations with the routine monitoring data, in part because of variations in outdoor pollution levels across a metropolitan area. These spatial variations are most pronounced for coarse particles and tend to be minimal for fine particles and sulfates. As a result, the actual exposure levels are poorly known, the contributions of outdoor air to these exposures can only be estimated, and important statistical difficulties result with respect to the epidemiology. These difficulties tend to obscure the true shapes of the dose-response functions, including the presence of thresholds, and have not been adequately addressed by

EPA or by the authors of the studies that EPA has selected.

As an example, air monitors in EPA's key study of acute mortality by particle size were located up to 60 km away from the population centers in three of the six cities involved. This makes it highly unlikely to find any effects attributed to coarse particles and other reactive (O₃) or local (CO) pollutants, since the measurements were made so far away from the bulk of the population.

4. Lack of plausible biological mechanisms. The problems of unknown exposures and of collinear pollutants make it difficult to reliably link specific pollutants with specific diseases, which is a prerequisite to formulating hypotheses for biological

mechanisms. In contrast to PM, this information exists for some of the gaseous pollutants, notably CO (heart disease) and O₃ (respiratory disease). Implausible combinations that have been reported include CO and respiratory disease, O₃ and heart disease, and sulfates and cancer. When the combinations of PM and O₃ are linked with respiratory disease, and a mechanism is known for O₃ but not for PM, it is difficult to accept the PM contribution as real, given the problems of measurement error discussed above. The same reasoning applies to the combination of PM and CO with respect to heart disease.

Technical Issues Pertaining to Long-Term Studies

1. Regional confounding by lifestyle, ethnic, and socioeconomic factors.

Long-term air pollution epidemiological studies depend on cross-sectional (spatial) contrasts, in which spatial patterns in health indicators are compared with spatial patterns in air quality in order to develop dose-response relationships. If either of the health or the air quality patterns were randomly distributed across the country, this procedure might provide useful results, but unfortunately, this is not the usual case. Air pollution patterns that have been reasonably stable over time include very high O₃ and PM levels in Southern California, moderate ozone levels across almost the entire region east of the Mississippi, and elevated PM_{2.5} and sulfur oxide levels in Appalachia and parts of the Northeast. Ischemic heart disease tends to be higher in the northeastern states, strokes are more common across a wide region of the South, respiratory disease mortality seems to be more common in certain mountain states, and people generally tend to live longer in Sunbelt cities and in the West. Further, some of these patterns have been changing over time, presumably due to diffusion of new medical procedures and changes in diet and lifestyle. Now, there are many reasons for these regional disease patterns that have nothing to do with air pollution, but, unless these exogenous factors are considered, a naive analysis will assign the regional differences in health indicators to air pollution. Examples of the non-pollution factors that should be considered include diet, lifestyle factors such as exercise, selective migration, and ethnicity. The two EPA studies considered none of these and the review shows that regional differences in physical activity could have accounted for almost all of the mortality gradients that one of these

studies assigned to air pollution, including fine particles and sulfates.

2. Inadequate consideration of all pollutants. None of the long-term studies have considered CO as a potential risk factor. One of EPA's key mortality studies considered only two pollutants, PM2.5 and SO42-, and thus provides no information on the role of particle size. The other mortality study had data on only six locations, so that the confidence limits on the relative effects of different pollutants were unacceptably wide. Neither study considered peak O3 levels as a risk factor (note that O3, SO42-, and PM2.5 tend to be collinear during hot weather and that there were major heat waves during the periods of study).

3. Latency periods between exposure and development of chronic disease.

EPA intends its annual standards to protect against chronic effects (they could also protect against the annual sum of acute effects, but in this interpretation, 24-hr standards would not be needed). Chronic effects, by definition, have their origins well before their existence is actually observed; typical latency periods run from 10-40 years. Neither of the long-term epidemiological studies selected by EPA considered prior exposures, which were considerably higher in the most polluted cities. Had they done so, the resulting dose-response functions would have more closely resembled those developed from acute studies. Thus, even if these two studies are taken at face value (which is problematic for other reasons, discussed below), they should not be interpreted as supporting the existence of chronic responses to air pollution. This alone calls into question the need for a separate annual standard for PM.

4. Investigation of pollutant thresholds. None of the long-term studies considered by EPA made an attempt to distinguish health effects in cities that violated the NAAQS from those in compliance. Two of the six cities in one mortality study exceeded the NAAQS for TSP during the study. Almost half of the cities in the more comprehensive mortality studies exceeded the NAAQS for TSP for the period of study, and the fraction of cities in violation for earlier periods would have been higher. Almost all major U.S. cities violated the O₃ NAAQS (and many still do). Given the problems in estimating true exposures and the collinearity among pollutants, the need for stricter air quality standards cannot be inferred from these studies.

Technical Issues Pertaining to Short-Term Studies

1. Misinterpretation of the apparent consistency of findings. EPA regards the apparent consistency of the many (over 30) studies of short-term PM effects throughout the U.S. and in other countries as a major element in its causal hypothesis. However, for each data set that has been reanalyzed by different investigators, agents (pollutants and/or weather factors) other than PM have been found to be equally plausible risk factors. In addition, in terms of mean effects on all-cause mortality, this consistency extends to all air pollutants (gases and particles) and to virtually all subsets of PM. This should be regarded, not

as support for causality, but instead as a warning that the apparent consistency may be derived from the common structure of the studies, rather than from physiological considerations per se. First, it is difficult to rationalize how such a wide range of air pollutants could all exhibit the same types of responses, let alone the same magnitudes, given the differences in personal characteristics and exposures to be expected in this wide variety of situations. Second, it is well known that "PM", even when classified by particle size, has very different characteristics in different parts of the country. Is it reasonable to expect exactly the same responses to wood smoke and sulfate particles? This is what EPA's claims of consistency implies.

3. Inadequate controls for seasonal and weather effects. The short-term epidemiological studies use temporal relationships between daily values of health indices and of air quality to postulate dose-response relationships. This requires accounting for all non-pollution factors that affect short-term health. Such factors include season (respiratory disease and mortality peak in the winter, in all climates), weather extremes (hot and cold), days of the week (more people die on Mondays; O₃ is lower on weekends but primary pollutants are higher), and holidays (mortality is higher on Christmas). Since the putative effects of air pollution are so weak (1-4%), it is necessary to account for even very small effects in the regression models. The studies in the literature vary considerably in their attention to these details; if days of the week are not accounted for, for example, (as sometimes has been the case) spurious relationships will result. Inadequate consideration of heat wave effects can result in overstating the effects of both O₃ and PM.

4. Improper consideration of lag effects. The only plausible mechanism for mortality occurring at the low air pollution concentrations we now enjoy is that of loss of homeostasis in severely compromised individuals (EPA has not advanced this hypothesis, however). In such cases, virtually any small environmental disturbance, including weather and pollution changes, might be sufficient to advance death by a few days. In such cases, a corresponding dip in the daily mortality counts should be seen on following days, and this has been in fact reported by several authors. In some cases, the net excess mortality over a week or so thus approaches zero. However, other authors have not reported this phenomenon but instead report only the combination of mortality delay (i.e., "lag") and pollutant averaging time that maximizes the apparent positive response. This is one of the reasons that reanalysis of a given data set by different investigators almost always produces different results.

A further consideration here is the physiological plausibility of either mortality or hospitalization occurring on the same day as an air pollution excursion. Clinicians and others have expressed doubt that such effects can happen, especially when one considers that most air pollutants are not uniform over the course of a day but exhibit well-defined diurnal peaks (O₃ in the afternoon or CO in the morning, for example). Heart attacks are most common in the morning, so that there may be a real probability of the

response preceding the cause.

Overall Conclusion

This review finds a lack of statistical and physiological support for the specific fine particle ambient air quality standards that EPA proposes. While the epidemiological studies do not exclude the existence of associations between fine particles and health responses, they fail to show responses that are specific to this class of particles and thus the need for a separate standard. Finally, EPA has relied on only part of the epidemiological literature and has neglected to consider several important factors bearing on the likelihood of causality. Given the fact that ambient air quality (including fine particles) will likely

continue to improve under current regulations, EPA could better serve the public interest by devoting additional resources to a comprehensive targeted program of research and monitoring.

Technical Comments* on EPA's Proposed Revisions to the National Ambient Air Quality Standard for Particulate Matter

prepared for the U.S. Department of Energy, Office of Fossil Energy (rev. 2-14-97)

1. Background

EPA's proposed revisions to the National Ambient Air Quality Standards (NAAQS) for particulate matter (PM) are based almost entirely on selected epidemiological studies that have been published in recent years. The important reference documents for EPA's proposal include: the revised criteria document for particulate matter (EPA/600/P-95/001cF, April 1996); the revised EPA Staff Paper (EPA-452/R-96-013, July 1996); the proposed regulations (Federal Register 61:65638, Dec. 13, 1996), and the draft regulatory impact analysis (no document number, dated December 1996). In these comments, the self-consistency of this body of information, key technical issues that arise and have been incompletely addressed, and new information that has since appeared in the scientific literature are examined. EPA proposed to retain the present standards for "inhalable" particles, i.e., those with aerodynamic diameters nominally less than 10 μm , i.e., PM₁₀, and slightly relaxed the 24-hr standard by allowing more exceedances per year. However, EPA proposed strict new standards for "fine" particles (those with aerodynamic diameters nominally less than 2.5 μm , i.e., PM_{2.5}) and this review emphasizes the smaller particles. The review begins with some introductory concepts that are needed to understand the scientific issues; reviews the studies that EPA selected as supporting their proposal, new studies that have since appeared, and existing studies that EPA should have considered. Then, some basic scientific issues that remain unresolved are discussed and the conclusions and recommendations of the review are presented.

1.1 Types of Particles in the Atmosphere

In general, fine particles are distinguished from larger particles by their sources and chemistry as well as by size. The first distinction is that particles formed in the atmosphere from gaseous precursors are initially quite small, mainly less than 0.1 μm , and then grow in size due to agglomeration and absorption of atmospheric water. These water soluble particles are often

mixtures of sulfate, nitrate, hydrogen (acid) and ammonium compounds and may be slightly acidic if they have not been in contact with atmospheric ammonia (gas). These particles can also be quite hygroscopic and tend to grow in size at high relative humidity, and thus tend to contain substantial particle-bound water. Organic particles may be formed as part of the photochemistry that produces ozone. Particles are formed from combustion processes that include volatilization of material at high temperatures followed by condensation in the atmosphere. These (insoluble) particles include soot and metal oxides. Health effects of some of these particles have been demonstrated in animal testing, at concentrations much higher than found in the ambient. Wood smoke is among those for which a biological mechanism has been demonstrated (Anon., 1995); breathing coal

smoke from domestic cooking has also been shown to have health effects (Xu and Wang, 1993). Transient health effects from acidic particles have only been demonstrated at concentrations about 2 orders of magnitude than those found in the ambient (EPA, 1996).

Particles larger than 2.5 μm include pollen, spores, fibers, other biological aerosols, tire erosion products, and small fragments of material from the earth's crust. These particles tend to be insoluble. Sources of crustal particles include road dust, material handling, mining, cement manufacturing, agriculture. Some of these (silica, asbestos, and biological materials) are also known to produce health effects. Shapes of these particles are also important (e.g., asbestos); soot is typically found in large irregular agglomerations with high surface-to-volume ratios, such that pollutant gases may readily be adsorbed. Soot may be found in both size fractions.

* Note that these comments incorporate material that was originally prepared for other purposes, some of which has been published in various forms. These publications include Section 12.4 of the PM Criteria Document and various publications by Lipfert and colleagues.

The arbitrary size cut at about 2.5 μm does not provide a complete separation of the two types of particles; the lower tail of the biological and crustal particle distributions will often extend into the "fine" particle collection, and some secondary water-soluble particles (notably nitrates) may exceed 2.5 μm .

It is important to realize that all air pollutants are defined for regulatory purposes by their officially-designated EPA "reference" measurement methods. For pollutant gases, these methods are generally quite specific to the actual chemical compounds (e.g., CO, SO₂, O₃, NO₂), but this is not the case for particles. The reference method for PM₁₀, for example, defines this quantity by the mass of material collected on a filter using a sampler meeting EPA specifications, without regard to the chemical compounds involved. This mass determination may be augmented or diminished by the processes of collection and analysis; volatile materials such as nitrates and organics may be lost and reactive gases may form compounds on the filters that were not present in the ambient. This latter process is most

important for nitrate and sulfate aerosols as collected on the glass fiber filters used in high-volume (TSP) samplers. Analysis has shown that this "artifact" sulfate is about 5 ug/m³ in the Eastern U.S., which is about half of the total SO₄²⁻ mass reported (Lipfert, 1994a). A large part of the typical difference between TSP and PM₁₀ is suspected to be artifactual, since mass is gained on the glass fiber TSP filters (SO₄²⁻ and NO₃⁻) and lost from the Teflon PM₁₀ filters (organics and NO₃⁻) (Lipfert, 1994a). Chemical-specific PM measurement methods include SO₄²⁻ and NO₃⁻ ions and carbonaceous particles. The latter are typically related to the darkness of filter paper stained by passing the air sample through; these determinations are referred to in various contexts as "smoke", "British smoke", coefficient of haze (COH), or (in California) "KM". The samplers

used for these determinations are thought to be selective to particles less than about 5-6 $\mu\text{g}/\text{m}^3$ in diameter, although the typically irregular shapes of soot particles may result in different lung deposition patterns than more spherical particles.

Finally, not all air pollutants are measured with current routine air sampling; potentially important omissions include aerosol acidity, elemental carbon, metals, aldehydes, and benzo(a)pyrene, which has recently been identified as a known human carcinogen. Sulfates and nitrates are measured routinely by only a few states.

1.2 Types of Epidemiological Studies Used to Support EPA's Proposal

All of the information on health effects associated with PM at current ambient levels comes from epidemiological studies. However, robust, specific pollutant-disease relationships have largely been lacking, in part because most of the studies have considered PM in general rather than specific types of particles. Further, no plausible biological mechanisms have been advanced in support of these various statistical associations. Like almost all environmental epidemiology studies, these studies are opportunistic in that they take advantage of data that were originally collected for other purposes. The dependent variables include daily counts of mortality and hospital admissions, annual mortality rates, diary studies on symptoms, and measures of lung function. The independent variables include data on individual and population characteristics and data from centrally located air pollution and weather instruments, but no information on individual exposures to pollution or to weather extremes. Also, only one study has considered the actual detailed chemical characteristics of the particles collected (Dockery et al., 1992). Further, the particle collection systems can create both positive and negative artifacts with respect to actual ambient conditions. Positive artifacts result from conversion of pollutant gases (such as SO_2) to particles (such as SO_4^{2-}) on the filter; negative artifacts result from loss of volatile material (such as nitrates and organics) after collection (Lipfert, 1994a). These uncertainties in exposure lead to serious difficulties in the statistical analyses that have been used to support the proposed standards (Lipfert and Wyzga, 1995a). These problems are discussed below in Section 1.3.

There are two types of observational epidemiological (statistical) studies in common use for studying air pollution. One type uses long-term contrasts among locations to derive statistical relationships between the air quality in different places and various health characteristics of the local population. This type is essentially a study of differences among places and is called a "cross-sectional" study. Studies of this type have been used to support annual air quality standards.

The second basic type of statistical study involves similarities in the timing of acute events, in which daily variations in air quality and weather are contrasted with daily variations in various health indicators. These time-series studies essentially

contrast population health statistics on the "bad" days with those on the "good" days, and the worst days usually have adverse weather conditions as well as poor air quality. Studies of this type have been used to support shorter-term ambient air quality standards such as 1-, 3-, or 24-hour standards.

Since failure to achieve either short- or long-term standards can have similar consequences in requiring reductions in emissions from specified sources, there are some interactions among the standards. However, large isolated sources with tall stacks (power plants, smelters) will tend to be more important for short-term concentrations, and a dense distribution of low-level area sources (vehicles, space heating) will tend to have the most effect on annual air quality standards (per unit of pollutant emission). These rules-of-thumb apply to emissions of primary pollutants; in contrast, secondary pollutants that are formed in the atmosphere from chemical reactions of secondary pollutants tend to affect larger areas relatively far from the primary sources and may be important at all time scales.

1.3 Exposure Error Issues Common to All Air-Pollution Epidemiological Studies

However, there have been several separate studies of personal exposures for a few individuals for periods up to 2 weeks (Lipfert and Wyzga, 1996)). Instead, data from outdoor centrally located air monitoring stations that are intended for routine environmental management and enforcement are used as surrogates for exposure to the complex mixtures that individuals receive in various indoor and outdoor situations. These data rarely (if ever) include speciation of particulate mass by compound or element, which makes it all the more difficult to postulate specific pollutant-disease relationships or to hypothesize biological mechanisms.

It must be emphasized that epidemiological studies in general cannot be used to "prove" cause and effect; their best use is in generating hypotheses to be tested in controlled settings. A major problem with observational epidemiology studies is the lack of data on individual exposures to air pollution. The use of surrogate exposure measures (such as ambient air quality data from central monitors) results in biased "dose-response" functions, such as a slope or regression coefficient (Pickles,

1982). If the error induced by using a surrogate is normally distributed with zero mean (in other words, when it creates random noise with no effect on the mean value), the slope is biased towards a shallower relationship, since the scale of exposures has been expanded due to the variability added from using a surrogate (Armstrong, 1990). This also biases the x-intercept (which might be regarded as a threshold) to the left (Lipfert and Wyzga, 1996). If the true dose-response function is non-linear (i.e., has curvature), this measurement noise will obscure this fact and make it appear to be linear (Yoshimura, 1990). Thus, in general, the types of exposure data that EPA has relied upon in setting NAAQS will tend to underestimate the risks of high exposures and overestimate the risks of low exposures. Note that people have respiratory defenses against most air

pollutants, and adverse health effects can occur in an individual only when these defenses have been overcome. This is why individual (rather than group-average) exposures are important and why it is reasonable to expect to find pollutant thresholds, even though the levels may vary among individuals.

Although detailed data on personal exposures are scarce for most pollutants, some generalizations can be made. Smaller particles settle out of the atmosphere more slowly and thus tend to have longer residence times and smoother spatial distributions. However, there are many sources of fine particles indoors that tend to obscure the relationships between actual personal exposures and the ambient monitoring data used in epidemiological studies. An important exception is sulfate aerosol (Dockery and Spengler, 1981), which is the only pollutant for which personal exposures correlate well with ambient measurements (probably because of the absence of indoor sources). Pollutants with especially poor relationships between personal exposures and ambient data include SO₂, CO, O₃, TSP. This is a result of the combined effects of indoor sources and sinks and of outdoor spatial variability due to local emission sources (Lipfert and Wyzga, 1996).

None of the "new" statistical studies have been able to firmly identify pollutant thresholds that might serve to support ambient standards, although only a few have made serious attempts. However, since all of them are based on surrogate exposure measures, this does not constitute evidence that such thresholds do not exist. A part of the failure to find thresholds results from the fact that the highest pollution days tend to not correspond with the worst health days; instead, such correspondence occurs more often at lower pollution values, nearer the mean. This could be interpreted either as a measurement error problem (perhaps people are not actually exposed to the highest pollution levels) or as a problem in the statistical analysis, but in any event this mismatching sounds a warning that a closer look should be taken at both possibilities. Knowledge of biological mechanisms are needed to aid in such interpretation.

Finally, when correlated pollutants have different exposure measurement errors, these differences will tend to control their relative performance in joint regression models that are intended to partition their effects (Lipfert and Wyzga, 1995a). The

pollutant with the least error will tend to dominate, assuming equivalent underlying toxicity. This means that the results of studies involving joint regressions of correlated pollutants with different estimated exposure reliabilities (which includes almost all of them) cannot necessarily be taken at face value (Lipfert and Wyzga, 1995a).

1.4 Bases for Comparing Epidemiological Studies

A review of this type will inevitably involve comparisons of associations involving different pollutants within a given study and between studies. Different authors use different paradigms for this purpose that are not always well defined. Common expressions in the literature include "consistent" or "stronger"

associations, or even "more significant." In this review, statistical significance is considered a threshold of credibility that must be overcome, taken here as a 5% probability level ($p < 0.05$, which requires a regression coefficient about twice its standard error). Significance depends on sample size and measurement error and should not necessarily be considered as a figure of merit; a study with $p < 0.001$ does not necessarily outrank one with $p < 0.01$, for example. "Mean effect" or elasticity at the mean is used as a quantitative, continuous figure of merit relating the endpoint to a risk factor; this measure is obtained by multiplying the regression coefficient by the mean value of the risk factor. If Poisson or log-linear models are used, the resulting mean effect is expressed as a fraction of the endpoint. If ordinary least squares regression is used, the mean effect is expressed as a fraction by dividing this product by the mean value of the endpoint variable. The relative risk at the mean pollution levels is obtained by adding 1.0 to the fractional mean effect.

Many authors express their results as relative risks, a measure that is probably most appropriate for specific individual health risks such as smoking or being a diabetic. It is less appropriate for air pollution, since we have no knowledge that the putative victims were actually exposed to the levels on which these figures are based. Further, authors differ with respect to the bases on which they express relative risks. Some authors express relative risks in terms of some arbitrary concentration level, such as 50 or 100 $\mu\text{g}/\text{m}^3$. This makes the risks due to small particles seem higher, since 50 $\mu\text{g}/\text{m}^3$ of PM_{2.5} is a much rarer situation than 50 $\mu\text{g}/\text{m}^3$ of PM₁₀, for example. Other studies have used the entire range (maximum - minimum) of the pollutant variable as a basis for risk comparison, or the "interquartile range" (IQR = 75th percentile - 25th percentile). This practice tends to make the pollutant with the most dispersed distribution appear to be the most harmful. For example, Thurston et al. (1994) reported that aerosol acidity posed the largest relative risk (1.5) for respiratory hospital admissions in Toronto, based on its range from minimum to maximum concentration. Ozone was stated to pose lower risks on this basis (1.34). However, the mean effect of ozone was about five times that of acidity, because high concentrations of acidity occur so much less frequently than high concentrations of ozone. Based on the results of Thurston et al. (1994), ozone should thus

be selected as the pollutant whose control should yield the most benefit (assuming no confounding due to temperature, etc.).

Also, the qualitative appearance of confidence limits for risk ratios depends on the risk levels selected for illustration; for example, a relative risk of 1.02 with confidence limits of 1.005 to 1.035 seems to be less impressive than one given as 1.2 with confidence limits of 1.05 to 1.35, yet these represent exactly the same situation, differing only in the exposure level selected to illustrate the risk. Use of mean effects has none of these problems.*

2. Key Epidemiological Studies in the EPA Proposal

2.1 Studies Intended to Support the Proposed Short-term (24-hr) NAAQS for PM_{2.5}

2.1.1 Mortality Studies. Although costs and benefits are not formally part of the consideration for revising the NAAQS, EPA has emphasized that their estimates of health benefits widely exceed their cost estimates. The major part of their estimated economic benefits comes from changes in mortality, in part because each life that is projected to be extended (note that no lives are actually "saved"[as reported by EPA]) is valued at several million dollars, regardless of the length of such life extensions. Epidemiological studies with mortality endpoints are thus very important in the analysis. However, only two studies specifically considered the association between daily mortality and fine particle concentrations, measured as PM_{2.5}, and their results are somewhat conflicting. Other studies have used TSP (particles up to 50 μ m), PM₁₀, COH, and KM. TSP and PM₁₀ are the most common PM metrics used in recent mortality studies.

Dockery et al. (1992) considered metropolitan St. Louis and a large area in Eastern Tennessee in the first attempts at time-series studies of particles classified by size and chemistry. These are two of the "Harvard Six Cities" that are discussed more fully below. Unfortunately, this effort was compromised by the short periods of record used and the large geographic areas that were required to obtain sufficient daily death counts for analysis. The St. Louis SMSA comprises eight counties, plus the city, for a total of almost 13,000 km². For eastern Tennessee, the geographic area was extended to include 11 counties (total area of about 12,000 km²).

A single monitoring site was used in each area, which recorded daily values for PM₁₀, PM_{2.5}, SO₄⁼, H⁺, SO₂, ozone, and NO₂. Carbon monoxide (CO) measurements were not included. Elemental composition data were also obtained from the particulate catch. The St. Louis site was located on the south side of the city in a residential area (Carondelet). The TSP data collected for use in a 1980 cross-sectional study (Lipfert, 1993) offer a means of comparing mean particulate values across the SMSA; in 1980, the SMSA-wide mean of 27 air monitoring stations was 84.2 μ g/m³, with a standard deviation of 17.2. The difference between the mean TSP values for the city and for the SMSA was statistically significant. Based on TSP data recorded at Carondelet in 1980

(80 ug/m³ [Dockery et al., 1989]), this site did not seem out of line with the rest of the SMSA on the basis of mean values; however, the question of interest here is the temporal correlation, which was not available.

Dockery et al. determined PM₁₀ by adding the fine and coarse particle mass determinations, thus capturing the measurement errors of both. In spite of this handicap, PM₁₀ was the only pollutant of the seven species to achieve statistical significance in St. Louis, although PM_{2.5} was close as was the coarse particle fraction. Thus, a major difference in mortality association by -----

* The selection of the most appropriate measure of health response also relates to the existence of thresholds. Using the

mean effect is tantamount to treating all concentration excursion the same, regardless of their absolute magnitudes. This "no threshold" paradigm is consistent with the fact that none of the published epidemiological studies has been able to establish the existence of such thresholds (measurement error may be partly responsible for this outcome). If, on the other hand, thresholds do exist, then it would be appropriate to consider risks specific to the higher concentrations, as was done, for example, by Abbey et al. (1991).

particle size was not indicated. One- and two-day lags were about equally effective. Neither SO_4^{2-} nor H^+ were even close to significance, nor were any of the gaseous species (SO_4^{2-} is a major component of the fine particle fraction). Elements normally associated with the coarse fraction were reported to have positive associations with mortality. The mean effects for PM_{10} and $\text{PM}_{2.5}$ were 0.041 and 0.030, respectively. The linearity plot showed that the first three PM_{10} quintiles all had about the same risk (1.0 - 1.01), with mean values from about 12 to 27 $\mu\text{g}/\text{m}^3$. The highest two quintiles both had an average risk of about 1.03 (note that these results could also be interpreted as a "hockey stick" dose-response function, with a threshold at about 25-30 $\mu\text{g}/\text{m}^3$ (24-hr average PM_{10})). The authors pointed out that the highest PM_{10} value in this data set was only 97 $\mu\text{g}/\text{m}^3$, well below the 24-hr NAAQS level of 150 $\mu\text{g}/\text{m}^3$. However, we have no assurance that levels are this low throughout the SMSA, nor do we know what TSP levels the decedents are actually exposed to. The estimated average TSP level for St. Louis for this time period was 64 $\mu\text{g}/\text{m}^3$, which is about 27% higher than the level implied by the PM_{10} data used in the study. Obviously, if the average PM level is 27% higher, there must be other locations in the metropolitan area with still higher PM levels, some of which could have been in violation of the NAAQS.

The portion of the study dealing with Eastern Tennessee was structured around a monitoring data set for Harriman, TN, which was similar to the St. Louis data. There was more topographical relief in the Tennessee study area, and almost half of the population lived in one county (Knox), which was not specifically represented by air quality monitoring data. None of the air pollutants was even close to statistical significance, although the regression coefficients were generally similar to those found in St. Louis. The mean effects were 0.048 for both PM_{10} and $\text{PM}_{2.5}$ and 0.07 for SO_4^{2-} . The linearity plot showed that the

second highest PM10 quintile had a risk of about 1.0, relative to the lowest quintile, while the remaining quintiles formed an approximate linear relationship. The estimated average TSP level for Knox County (the major population center of the area under study) for this time period was about 55 ug/m³. The lack of significance shown for Eastern Tennessee may have resulted from the smaller average number of daily deaths (15.5 vs. 56 for the St. Louis area). If the algorithm of Ito et al. (1996) is used to adjust the standard errors to an equivalent daily death rate, both PM10 and PM2.5 would have achieved significance, with about the same regression coefficients as in St. Louis. However, SO₄⁼, H⁺, SO₂, NO₂, and O₃ would have not have achieved statistical significance. This implies that the non-sulfate particles must have been more important.

Another important contribution of Dockery et al. is the testing of chemical constituents of PM for associations with daily mortality. Positive associations were reported in both metropolitan areas for Al, Ca, Cr, Fe, and SiO₂, all of which are commonly found in particles associated with the earth's crust (rather than with combustion products, per se). Crustal particles were found to be associated with respiratory hospital admissions in Anchorage, Alaska, for example (Gordian et al., 1996 (discussed in more detail below)).

This study was followed by an extended analysis of all six cities, covering the period 1976 to 1987 (Schwartz et al., 1996). Although the authors assigned the entire acute mortality effect to "combustion-related" fine particles (even though such particles had not been specifically addressed), this outcome was largely preordained by the design of the study. As in the previous study, the extended analysis was based on data from a single dichotomous sampler located in each of the six areas (and not always in the population center), while the mortality data being fit to these data were taken from surrounding areas sometimes comprising as many as 9 counties and areas of up to 150 km in extent. This design provides much more reliable estimates of exposure for fine particles than for coarse particles, by virtue of the more precise and accurate measurements of PM_{2.5} from the dichotomous samplers used* and the much more uniform spatial distributions of fine particles across each of the six metropolitan areas. Since random errors in exposure tend to bias the regression slope towards the null and towards nonsignificance, the Six City study design contains a strong bias in favor of fine particles. This bias is discussed in detail by Lipfert and Wyzga (1997). The resulting increased measurement error for coarse particles will have biased their regression result towards non-significance, and this fact alone could explain the differences reported by Schwartz et al. Further, it is necessary to consider gaseous pollutants (such as CO or O₃) simultaneously with particles to provide robust estimates of the relative responsibilities for the implied daily mortality gradients.

Finally, the model specifications used by Schwartz et al. may be sensitive to differences in pollutant frequency distribution characteristics according to particle size. For example, during

most of their study, size-classified particles were collected only every other day. Schwartz et al. elected to use a 2-day average of PM concentrations in their regression analysis, which amounts to assuming a 2-day persistence for air quality. Since smaller particles tend to have longer residence times in the atmosphere, this assumption is more realistic for fine than for coarse particles and thus creates an additional statistical bias in favor of PM_{2.5}.

The authors' conclusions in this paper are thus not supported by their data and analyses. As concluded by Lipfert and Wyzga (1997), a more appropriate conclusion would have been that it was incapable of determining differences in association with mortality by particle size. In addition, their failure to

consider gaseous pollutants raises questions as to whether the mortality associations can be realistically assigned to any PM constituent.

2.1.2 Hospitalization Studies. Thurston et al. (1994) examined associations between daily admissions to 22 acute care hospitals in metropolitan Toronto and various air pollution concentrations monitored daily at a central site for 6 weeks during July and August for 1986-88. The end points were admissions for respiratory causes (all respiratory and asthma) and for selected "control" causes. These data were apparently for all ages. Air quality data included specially collected size-classified particle concentrations and routine data on TSP, O₃, SO₂, and NO₂. PM₁₀ was obtained by summing the fine and coarse fractions of the dichotomous samplers and is thus affected by errors in both measurements. Thurston et al. noted that 1988 had a very hot summer and higher pollution levels. Ozone was highly correlated with temperature, as were SO₂-, NO₂, PM₁₀, and the coarse particle fraction of PM₁₀ (but not PM_{2.5}); these correlations were based on data from which seasonal trends and day-of-week effects had been removed. The regressions model used trigonometric functions to control for seasonal variability; this has the disadvantage of treating all years the same. Since 1988 was clearly different, this model may have not controlled adequately for weather effects. Humidity was not considered as a potential confounder and no special treatment was given to days of peak temperature and humidity. Such days have been shown to be risk factors for mortality in the Northeastern U.S., for example (Schwartz and Dockery, 1992).

Thurston et al. presented their regression results in terms of p values and regression coefficients; for the purposes of this review, the latter were converted into mean effects, as discussed above. All of their models included daily maximum temperature and either one or two air pollutants. In the models that included only one air pollutant, the mean effects on total respiratory admissions were, in decreasing order: O₃, PM₁₀, coarse fraction of PM₁₀, TSP, PM_{2.5}, SO₂-, TSP - PM₁₀, and H⁺. However, Thurston et al. chose to rank in terms of statistical significance, which places acidity first, instead of ozone. Similar results were obtained for asthma admissions. Ranking by mean effect ranking strongly suggests that fine particles are not the most important pollutants in Toronto in terms of their effect

on respiratory hospital admissions. For example, Thurston et al. (1994) reported that aerosol acidity posed the largest relative risk (1.5) for respiratory hospital admissions in Toronto, based on its range from minimum to maximum concentration. Ozone was stated to pose lower risks on this basis (1.34). However, the mean effect of ozone was about five times that of acidity, because high concentrations of acidity occur so much less frequently than high concentrations of ozone. On this basis, we conclude that ozone is the pollutant of most concern in this study of Toronto hospital admissions.

* The samplers used allocated only 10% of the sampled air stream to the coarse particle catch, which greatly reduces the amount of mass available for analysis. In addition, there was a problem

with loss of coarse particles during shipment from the field sites to the analytical laboratory (Spengler et al., 1986; Dzuby and Barbour, 1983). These two factors result in a measurement precision at least twice as good for fine particles, compared to coarse particles.

In two-pollutant models, the combined mean effects were about the same for all pollutants (as has been noted by Lipfert and Wyzga (1995b) for mortality studies). In models including O₃ with PM_{2.5} or SO₄²⁻, O₃ always had the larger mean effect. However, ozone could have been confounded by hot day effects. Detailed results were not presented for the "control" admissions, but none of the cross-correlations between control admissions and the air pollution variables were significant.

2.1.3 Other Morbidity Effects. No other acute morbidity studies specifically involving fine particles were cited in the Decision Document.

2.1.4 Summary of Acute Health Effects by Type of Particle. Two related acute mortality studies considered size-classified particles. Both of them were compromised by substantially increased measurement errors for the coarse particles that are included in PM₁₀. In spite of this handicap, PM₁₀ had about the same or larger effects on daily mortality than either PM_{2.5} or SO₄²⁻. For daily respiratory hospital admissions, PM₁₀ had a larger mean effect than PM_{2.5} in Toronto, and in a study not specifically considered by EPA (Gordian et al., 1996), crustal particles (PM₁₀) were specifically shown to predict daily respiratory outpatient visits. Sulfate had lower mean effects on respiratory hospital admissions than non-specific particles and was not significant in one of the acute mortality studies.

2.1.5 Conclusions on Short-term Supporting Studies. The studies used by EPA in support of their proposed 24-hr NAAQS for PM_{2.5} do not find effects that can clearly be attributed to PM_{2.5}.

Instead, associations with acute health effects do not seem to be specific to any single type of particle or gas (Lipfert and Wyzga, 1995b). There appears to be as much evidence implicating PM₁₀ or the coarse fraction of PM₁₀ as there is for PM_{2.5}. There is no evidence specific to SO₄²⁻ suggesting that those particles are any more toxic than other PM constituents; instead, given its statistical advantage resulting from its much smaller exposure error, the evidence suggests that SO₄²⁻ must be less toxic than

other constituents of PM.

2.2 Studies Intended to Support the Proposed Annual NAAQS for PM_{2.5}

Although several cross-sectional mortality studies involving long-term mortality were discussed in detail in the Criteria Document (Section 12.4), EPA focused on only two of them in their Decision Document. These were recently published prospective cohort studies, which have the advantage of dealing with individuals rather than community-wide averages, but are still limited by the use of community based air quality monitoring data, rather than individual exposures.

2.2.1 Long-Term Mortality Studies. EPA only considered two long-term mortality studies, which are critiqued below. Studies that were not considered by EPA include the prospective study by Abbey et al. (1991) of California Seventh-Day Adventists that failed to find mortality effects due to any measure of PM, and the population based study of Lipfert (1993) that found equivocal effects due to TSP (but no significant associations with PM₁₅, PM_{2.5}, or SO₄₂-). These omitted studies are considered further below (Section 3).

2.2.1.1 The Harvard Six Cities Study. Dockery et al. (1993) analyzed survival probabilities among 8111 adults who were first recruited in the mid-1970s in six cities in the eastern portion of the United States. The cities are: Portage, WI, a small town north of Madison; Topeka, KS; a geographically-defined section of St. Louis, MO; Steubenville, OH, an industrial community near the West Virginia-Pennsylvania border; Watertown, MA, a suburb of Boston; and Kingston-Harriman, TN, two small towns southwest of Knoxville. This selection of locations thus comprises a transect across the Northeastern and Northcentral United States, from suburban Boston, through Appalachia, and into the upper Midwest and Great Plains.

The adults were white and aged 25 to 74 at enrollment. Follow-up periods ranged from 14 to 16 years, during which from 13 to 22% of the enrollees died. Of the 1430 death certificates, 98% were located, including those for persons who had moved away and died elsewhere. However, no information was given in the paper about the actual locations of death. The bulk of the analysis was based on all-cause mortality; no mention was made of subtracting external causes of death. The Cox proportional hazards model was used to estimate coefficients for the individual risk factors after stratifying by sex and age (5-year groups).

Air monitoring data were obtained from routine sampling stations and from special instruments set up by the research team. The air pollution exposure data used in this analysis were based on data from a central station in each community rather than on individual exposures. Steubenville, Kingston-Harriman, and St. Louis were the most polluted (and also had the oldest and least educated cohorts and the heaviest rates of smoking) among the six cities.

Most of the air quality measures were averaged over the period of study, in an effort to study long-term (chronic) responses; the specific averaging periods varied by pollutant. However, the size-classified particulate data began in 1980 while TSP data began in 1974; from 1974 to 1980 there were large reductions in TSP (and probably in the size-classified particles as well), so that it appears that the size-classified data may be less representative of cumulative exposures than TSP. Sulfate appeared to be intermediate in this regard. Exposures prior to the study (which were much higher in the dirtiest cities and often violated the NAAQS) were not considered. In this sense, there is a mismatch in time between the air quality data, which were obtained after the study began, and the descriptive data on individuals, which pertain to the period before entry into the

study.

The effects of air pollution were evaluated in two ways: by evaluating the risks of residence in each city relative to Portage (the city with the lowest pollution levels for most indices), and by including air quality levels directly in the regression models. Since only six different values were available for each pollutant, the effective degrees of freedom are greatly reduced by this procedure, relative to the other variables and to what could have been done had personal exposure data been available.

Individual characteristics of the members (and thus of the decedents) considered included smoking habits, an index of occupational exposure, body mass index, and completion of a high-school education. The index of smoking rate used here was pack-years, defined as the average number of packs of cigarettes smoked per day times the number of years of smoking. This metric is also a function of age. Current and former smokers were treated separately. The index of socioeconomic status used was having less than a high school education; Rogot et al. (1992) show that this index is a good measure of mortality differences due to differences in education for white men, but not for white women. For women, there is a more continuous relationship between (reduced) relative mortality risk and increased levels of educational attainment.

The index of occupational exposure to air pollution (dusts or fumes) did not take into account the nature of the agents involved or the details of exposure. Occupational exposure to dusts or fumes was not found to be a significant risk factor; this outcome may have resulted from the lack of specificity of the index used. The average percentages of subjects having occupational exposure were high, ranging from 28 to 53%, with an average across all cities of 45%.

The index of physiology used was the body mass index (BMI), defined as weight divided by height squared (kg/m^2), treated as a linear relationship. The relative risk (RR) of increased body mass was similar to that found by Sandvik et al. (1993) (where it was not statistically significant), but other investigators have found that the relationship is U-shaped rather than linear and that it may interact with other risk factors, especially smoking

(Gronbaek et al., 1994). The nonlinear relationships imply increased mortality risk for those at either end of the scale, either obese or very thin.

In multi-factor regressions, the effect of inadequate representation of a confounder is to inflate the importance of other variables. In this analysis, for example, the excess mortality risk in Steubenville was reduced from 67% to 26% by adjusting for age, smoking status, education, and body-mass index. It is likely that the adjustments would have further reduced the "unexplained" risk if more detailed attention had been given to each of these factors. In addition, a number of potential confounders were not considered.

No consideration was given to possible independent effects of occupation classification, other personal lifestyle variables such as diet or physical activity, migration, or income. Presumably, each subject was characterized by his status at entry to the study; follow-up data on possible changes in risk factors over time (such as smoking status, occupational exposure, or body mass) were not mentioned. Figure 1 (from Lipfert, 1995) shows how differences in the fraction of the elderly population with sedentary lifestyles in each area (these data were not available for Topeka) can account for most of the mortality gradient, and how the slope of that relationship matches that found in an independent study in California (Breslow and Enstrom, 1980). The figure shows that differences in lifestyle among the six areas could explain most of the mortality risk that was attributed to air pollution.

It should also be noted that new independent studies (Mesink et al. 1996) continue to identify lack of physical activity as an important risk factor for heart disease and premature mortality. Also, new studies (Brody, 1996) have found that improper maternal diet can predispose an infant towards heart disease later in life; it is not unlikely that the Appalachian region of the U.S. that is now high in SO₄= aerosols (and PM_{2.5}) could have been affected by poor nutrition during the depression years when today's adults were in utero.

Statewide mortality data indicate substantial differences in survival rates across this transect of the Northeastern U.S that are similar to those that were observed among the six cities; this suggests that the responsible factors constitute regional trends. Dockery et al. report a long-term average mortality rate of 16.2 deaths per 1000 person-years in Steubenville and a rate of 9.7 in Topeka, yielding a range in average (crude) relative risk of 67% among the six cities. After individual adjustment for age, smoking status, education, and body-mass index, the range in average relative risk was reduced to 26%. In comparing the most and least polluted cities, Dockery et al. also report elevated risks for cardiopulmonary (ICD-9 401-440, 460-519) causes (1.37, [1.11-1.68]) and lung cancer (ICD-9 162) (1.37, [0.81-2.31], not significant). The relative risk for all other causes of death was 1.01 (0.79-1.30).

Dockery et al. report that "mortality was more strongly

associated with the levels of fine, inhalable, and sulfate particles" than with the other pollutants, which they attributed primarily to factors of particle size. They provided relative risk estimates and confidence limits based on the differences between air quality in Steubenville and in Portage for these three pollutants. However, it is a relatively simple matter to independently estimate these coefficients from the adjusted risks and pollutant levels reported for each of the six communities; these estimates are given in Table 1 (from Lipfert, 1995) as a means of comparing the various pollutants and combination of pollutants. Note the absence of information on CO as a risk factor. The data for 1970 TSP (lagged about 12 years) were obtained from Lipfert (1978), assuming that Madison could represent Portage, WI. The estimates made by this procedure

shown in Table 1 correspond quite closely to the figures given by Dockery et al. based on output from the Cox proportional hazards regression model. However, because there are only 6 degrees of freedom for the air quality data, the confidence limits in Table 1 are somewhat wider than those for the risk factors having individual data. As in the original paper, the relative risks are based on differences in average air pollution levels between Steubenville and Portage.

Table 1 Estimated Relative Risks in Six U.S. Cities
Associated with a Range of Air Pollutants (from Lipfert, 1995)

species	pollution		range*
	regr. coeff.	std error	
rel. risk 95% CIs (n=6)			
PM15	0.0085	(0.0026)	28.3 1.27
(1.04-1.56)			
PM2.5	0.0127	(0.0034)	18.6 1.27
(1.06-1.51)			
SO42-	0.0297	(0.0081)	8.5 1.29
(1.06-1.56)			
TSP	0.0037	(0.0014)	55.8 1.22
(0.99-1.53)			
1970 TSP	0.0014	(0.00044)	154 1.25
(1.03-1.50)			
TSP-PM15	0.0042	(0.0032)	27.5 1.12
(0.88-1.43)			
PM15-PM2.5	0.0178	(0.0098)	9.7 1.19
(0.91-1.55)			
PM2.5-SO4	0.0255	(0.0029)	8.4 1.24
(1.16-1.32)			
PM15-SO4	0.0121	(0.0034)	18.1 1.24
(1.05-1.48)			
SO2	0.0093	(0.0032)	19.8 1.20
(1.01-1.43)			
NO2	0.0126	(0.0046)	15.8 1.22
(1.00-1.49)			

*range = (maximum value - minimum value) for the 6 cities

Dockery et al. noted that ozone levels varied little among the six cities; this might not have been the case if a measure of peak concentration had been used for comparison instead of the annual means. No relationship was found for aerosol acidity (H+), but only limited data were available. Table 1 shows only small differences in the relative risks among many different pollutants, owing in part to the strong collinearity present in this data set. TSP and the coarse particle variables created through subtracting PM15 from TSP and PM2.5 from PM15 were not significant, but the non-sulfate portion of PM2.5 had the tightest confidence limits (to create this variable, SO42- was

multiplied by 1.2, assuming an average composition of NH_4HSO_4 , before subtraction). Also, the estimated 1970 TSP variable performed better than the TSP data used by Dockery et al. (ca. 1982). However, all of the differences in relative risks among pollutants and their confidence limits could have occurred due to chance, given only 6 observations.

Among the six cities, Steubenville appears to be the most important. It was the only city for which the residual risk, after accounting for age, race, sex, smoking, education, body mass, and occupational exposure (however crudely), was significantly different from zero. It is also the only location on Figure 1 with incremental mortality risks not explained by increased frequency of sedentary lifestyle. Steubenville has

traditionally been a city of heavy industry, primarily metalworking, and it is logical to assume that the chemical characteristics of its air pollution differ from more heterogeneous urban areas such as St. Louis, Topeka, or Watertown (Boston). For example, monthly average TSP readings as high as 800 ug/m³ were recorded in Steubenville during the period before the study (Lipfert and Wyzga, 1991) and airborne Fe and Mn concentrations there were among the highest in the nation ca. 1980 (Lipfert et al., 1988).

Comparison of the pollution risks among the various available cohort subsets is one of the most important outcomes of a study on individuals. Such comparisons must account for the higher variability among subgroups, however, and Dockery et al. were not capable of distinguishing differences in excess risks between subgroups less than about 18% (i.e., a relative risk of 1.18 cannot be distinguished from one of 1.36, for example). Although none of these subgroup differences were statistically significant, the mortality risks associated with area of residence (and thus air pollution) were higher for females and for smokers and the risks were also higher for those occupationally exposed compared to the nonexposed. However, the personal exposures of nonsmokers and of those without occupational exposures are more likely to correlate with the outdoor monitoring data used here than those with these additional sources of exposure. It is possible that the relative risk estimates for outdoor air pollution for these lesser exposed subgroups (1.19 and 1.17, respectively) might be the most reliable estimates in the study. However, risks for these subgroups are just as likely to have been overstated by lifestyle confounding as for the others.

Dockery et al. were quite cautious in their conclusions, stating only that the results suggest that fine-particulate air pollution "contributes to excess mortality in certain U.S. cities." However, EPA has moved from this "suggestion" to a firm conclusion, which they used as a basis for new regulation.

2.2.1.2 The American Cancer Society Study. Pope et al. (1995) analyzed 7-year survival data (1982-89) for about 550,000 adult volunteers obtained by the American Cancer Society (ACS). This cohort is not a random sample of the U.S. population; it is 94% white and better educated than the general public, with a lower

percentage of smokers than in the Six City Study. The (crude) death rate during the 7.25 years of follow-up was just under 1% per year, which is about 20% lower than expected for the white population of the U.S. in 1985, at the average age reported. In contrast, the corresponding mortality rates for the Six-Cities study discussed above tended to be higher than the U.S. average. No mention was made of residence histories for the decedents; matching was done on residence location at entry to the study. Causes of death considered included all causes, cardiopulmonary causes, lung cancer, and all other causes.

The Cox proportional hazards model was used to define individual risk factors for age, sex, race, smoking (including passive smoke exposure), occupational exposure, alcohol consumption, education,

and body-mass index. The deaths, about 39,000 in all, were assigned to geographic locations using the 3-digit zip codes listed at enrollment into the ACS study in 1982. Relative risks were then computed for 151 metropolitan areas defined by these zip codes and were compared to the corresponding air quality data, ca. 1980. The sources of air quality data used were the EPA AIRS system for sulfates, as obtained from high-volume sampler filters for 1980, and the Inhalable Particulate Network for fine particles (PM_{2.5}). The latter data were obtained from dichotomous samplers during 1979-81; Pope et al. used the values from this data base reported by Lipfert et al. (1988) but only 50 PM_{2.5} locations could be matched with the death data. The correlation between the two pollutants was 0.73. No other pollutants were considered. The authors assumed that 1979-81 pollution values would be representative of long-term cumulative exposures, in keeping with the objective of analyzing chronic effects. The sulfate values from the inhalable particle filters, which are thought to be free from artifacts, were not used by Pope et al. As in the Six City Study, all estimates of air pollution exposure were based on data obtained from central monitors in each city, rather than from personal exposures.

Pope et al. took great care with the potential confounding factors for which data were available. Several different measures of active smoking were considered, as was the time exposed to passive smoke (the only study to do so). The occupational exposure variable was specific to (any of) asbestos, chemicals/solvents, coal or stone dusts, coal tar/pitch/asphalt, diesel exhaust, or formaldehyde. The education variable was an indicator for having less than a high-school education. However, alcohol use and body-mass index were considered as linear predictors of survival, whereas other studies have indicated these effects to be non-linear (U- or J-shaped) (Gronbaek et al., 1994; Lew and Garfinkel, 1979; Doll et al., 1994). Risk factors not considered by Pope et al. include migration, income, employment status, dietary factors, drinking water hardness and physical activity levels, all of which have been shown to affect longevity (Pocock et al., 1980; Belloc, 1973; Sorlie and Rogot, 1990). Pollution coefficients were reduced by 10-15% when variables for climate extremes were added to the model.

Pope et al. found very consistent pollution risks for all-cause mortality for males and females and for ever-smokers and

never-smokers. The relative risks for air pollution were slightly higher for females for cardiopulmonary causes of death. The lung cancer-sulfate association was only significant for males, except for male never-smokers. The adjusted total mortality risk ratios (computed for the range of the pollution variables) were 1.15 (95% CI = 1.09-1.22) for sulfates and 1.17 (95% CI = 1.09-1.26) for PM_{2.5}. When expressed as log-linear regression coefficients, both pollution measures were quite similar: 0.0070 (0.0014) per ug/m³ for SO₄²⁻ and 0.0064 (0.0015) for PM_{2.5}; however, it is possible that the SO₄²⁻ results have been biased high by the presence of filter artifacts (Lipfert, 1994a). Expressed as the percentage of mortality associated with air pollution at the mean values and corrected for SO₄²⁻ filter artifact, Pope et al. found mean effects of about 5% for sulfate

and 12% for PM_{2.5}. No significant excess mortality for the "other" causes of death was attributed to air pollution.

In some ways, the ACS study was an improvement on the Six Cities study, because of its greater care in representing some of the confounders and its increased geographic scope. However, it suffered from omitting many of the same potential regional confounders and its use of only two highly related air pollutants. Other types of particulate matter and carbon monoxide were not investigated.

2.2.2 Other Long-term Health Effects. There are no long-term studies of air pollution and hospitalization, in part because utilization of hospital facilities is strongly influenced by supply and demand factors (Lipfert, 1994b). This imposes a severe limitation on examining coherence among long-term health endpoints.

The EPA decision document referred to Table 13-5 of the CD for other morbidity evidence supporting the need for an annual PM_{2.5} standard. In that Table, only the 24-Cities Study (discussed in detail below) actually considered fine particles in this regard. The previous Six City study was also included, but its comparisons were limited to TSP and PM₁₅/PM₁₀. Data from Abbey et al. (1995) for SO₄²⁻ in California were also included with respect to bronchitis in children. Using EPA's comparison criteria for relative risks (based on arbitrary increments of air quality), the "coarse" particles (TSP, PM₁₅ or PM₁₀) had an average odds ratio for children's bronchitis of 3.12 with a standard error of 0.16 (between studies), while the "fine" particles (H⁺, SO₄²⁻, PM_{2.5}) had an average odds ratio of only 2.26, with a standard error of 0.36. Thus, the smaller particles do not show greater toxicity for this endpoint. For decreased lung function in children, there were no significant differences across the Six Cities, and the decrements in the 24 Cities were not statistically different among any of the particle measures considered (H⁺, SO₄²⁻, PM_{2.5}, PM₁₀). Note that comparisons of respiratory effects due to ozone were not made in CD Table 13-5. The Six and Twentyfour Cities studies considered bronchitis prevalence, while the study of Abbey et al. (1995) considered incidence of new cases. The odds ratio for incidence was lower (1.39), but data for only one pollutant (SO₄²⁻) were reported by EPA.

2.2.2.1 The 24-Cities Study of Children's Respiratory Health. In response to concerns about the toxicity of acid aerosols, a cross-sectional analysis of the respiratory health of children (ages 8-10) in 24 small towns in the U.S. and Canada was begun in 1988 and continued through 1991 (Dockery et al., 1996; Raizenne et al., 1996; Spengler et al., 1996; Brook and Spengler, 1996). The protocol involved monitoring the local air quality in each town for about a year for a wide range of species, obtaining parental responses in September and October on children's respiratory health through questionnaires, and testing the pulmonary function of the children from October to May. Roughly one third of the communities were investigated in each of the 3 school years involved. The authors of these studies reported

that symptoms of bronchitis were associated with acidity and sulfates and that reduced pulmonary function was associated with particle acidity. These results were based on about 13,000 to 15,000 children, but the air quality measures were obtained from only 24 central monitoring stations. No individual exposures to air pollution were considered.

Reanalysis of the published data suggests that other interpretations may be equally or more likely. First, the design of the study has certain shortcomings. Since data were not taken simultaneously in all communities, both time and space are variables in the study. For example, 1988 was a high ozone and acidity year in much of the U.S. (Thurston et al., 1994), and it is possible that the monitoring data obtained for some of the 8 towns studied that year were not representative of long-term exposures. However, Brook and Spengler considered this problem for some (but not all) of the communities and dismissed it for the aggregate. In addition, lung function testing was performed sequentially throughout the school year and it is possible that seasonal variations may have biased the results for specific communities. The data of Spodnik et al. (1966) and of McKerrow and Rossiter (1968) show seasonal swings of a few percent, which is of the same order of magnitude as the air pollution effects reported by Raizenne et al. (1996). No mention was made of accounting for acute lung function responses to air quality differences on the days of testing. A priori, transient acute responses would be expected to exceed differences in long-term averages.

The results for respiratory symptoms in the 24 cities (Dockery et al., 1996) show that only 5 of 64 combinations of symptoms categories and pollutants were statistically significant; although not all of these comparisons are independent, this is close to what might have been expected just due to chance. In addition, the stratified sample analysis (Figure 2 of Dockery et al.) shows that the effect of acidity on bronchitis symptoms depends almost completely on the contrast between Eastern U.S. and Canadian cities. When U.S. cities were considered alone, the odds ratio was 1.0, and when only eastern cities were considered, the odds ratio exceeded 1.0 but was not significant. Thus, the acidity relationship is not robust. Also, only the first year of study (the high ozone year) was significant when the data were stratified by year. Presence of molds in the home was not

mentioned as a potential (confounding) risk factor (as determined separately by Spengler et al. (1994) for the same children and by Dales et al. (1991) for Canadian children), yet use of a humidifier was shown to define the most sensitive subgroup. Regression analysis of the tabulated aggregated data for each city (n=24) did not produce plausible relationships (smoking indicated a negative effect on bronchitis prevalence, for example). It is possible that the use of parental recall to define the endpoints produced too much noise for credible findings; it is also possible that there are systematic differences between the U.S. and Canada, either involving climate or definitions or diagnoses of symptoms. All of these factors suggest that there are hypotheses that provide more reasonable alternatives to the authors' conclusion that aerosol acidity was

the causal agent.

The results for lung function were more amenable to reanalysis. Figure 2 (of this review) plots the city-specific data for the average forced vital capacity (FVC, a measure of lung volume), the fraction of children who had less than 85% of the FVC predicted for that community (low FVC), after adjustment for age, sex, height, and weight, and the peak expiratory flow rates (PEFR). The data points are clustered geographically according to the authors' stratification of locations: the "SO₄ belt" represents Appalachia; the "transport region" is generally downwind of these locations; west coast stations are in California, including one in a high ozone region; "background" stations are in Arkansas, South Dakota, Saskatchewan, and British Columbia. When these plots are viewed without regard to geographic clustering, they appear to support the hypothesis that acidity degrades lung function. However, for FVC, the variation within the "SO₄ belt" is so large and unrelated to acidity that a regional bias factor is suggested, as an alternative to the causal acidity hypothesis. To evaluate this alternative hypothesis, the 22 aggregated observations (two of the towns refused to participate in this phase of the project) were regressed against air quality, regional dummy variables, and potential confounders such as parental smoking and education level. These regressions showed that the regional dummy variable was a more powerful predictor of FVC than any air pollutant (by deleting the dummy variables, this regression technique also reproduced the result for acidity reported by Raizenne et al. on the basis of individuals). For example, the best regression for FVC included the SO₄-belt regional dummy variable (-) and education (+) (and no air pollutants) with an R² of 0.64; when regional dummy variables were excluded, the best regression included acidity and peak ozone and had an R² of 0.56. The conclusion thus follows that geography per se (i.e., long-standing regional differences) is a better predictor of these small lung function differences than air pollution. The implied mean effect of acidity on FVC was about 1.6%; of peak O₃, about 1.9%. However, for peak flow, acidity was a better predictor than the regional dummy variable (mean effect, about 4%). These increments represent very modest effects on lung function.

The lung function results should be interpreted as suggesting

that Appalachian children tend to have smaller lungs (as indexed by FVC) than children elsewhere in the U.S. and Southern Canada. This might be a result of chronic exposure to some type of air pollution, but that hypothesis seems no more likely than one related to regional ethnicity, diet or exercise, given the known regional differences in these factors (Lipfert, 1995). Particulate acidity as a long-term risk factor is an especially problematic rationale, given the low probability that ambient acidity as measured at central stations would actually be delivered to the lungs. First, ammonia in the breath will tend to neutralize inhaled acids, especially for those typically smaller particles that might be found near sources. Second, acid particles that penetrate indoors will tend to be neutralized by indoor ammonia, coming from pets, household cleansers, or

personal sources. Acidity is perhaps the only air pollutant against which we have two layers of defenses, and it is difficult to rationalize its role as a cause of long-term health effects. Ozone may well be a better candidate for changes in lung function, especially since ozone has so much larger effect on acute lung function changes than acidity (EPRI, 1994).

It is also of interest to compare the results of the 24-City Study with those of its predecessor, the Harvard Six City Study. In that study, "bronchitis" was associated with TSP and with PM15, and later with acidity for a subset of the cities. However, for lung function for adult never-smokers, Steubenville (the most polluted city) had a higher average lung function than would have been expected for that population, in spite of the much worse air quality that was experienced there during the 1960s and 1970s. For children, using TSP as a predictor, both St. Louis and Steubenville had lung function averages higher than expected (see Lipfert [1994b], pp. 471-3 for plots of these data). Thus, the Six City data seem to suggest that worse air pollution is accompanied by better lung function! In the Six City Study, the entire range of TSP was associated with about 3% in lung function, which seems to be a smaller incremental effect than is now seen in the 24 cities, even though pollution levels are currently much lower. Lipfert (1994b) also noted that use of long-term average ozone as a health predictor obscures the effects of peak ozone excursions; he estimated that differences in peak ozone could have predicted the Six City lung function results, with a mean effect of about 3.6%.

2.2.3. Conclusions from Long-Term Supporting Studies. While the prospective cohort mortality studies broke new ground in terms of cross-sectional studies, they suffered from severe data limitations, in large part because they were not designed for this purpose from the outset and thus the necessary data on each subject were not collected at entry to the studies. The regional confounding between sulfate and/or fine particles that exists in the U.S. precludes valid estimates of their long-term effects on this basis. Potential confounders include lifestyle, diet, economic and employment status, and migration. These studies thus provide no robust basis on which to estimate the existence of long-term health effects. This also precludes the use of these studies to conclude that acute mortality effects are not dominated by mortality "displacement" (subsequent decreases in

daily deaths that compensate for the increases associated with the environment) or as a means of estimating the degree of life-shortening in the acute mortality studies. Failure to consider previous or cumulative exposures to air pollution overstates the effects attributed to current exposures. The facts that peak ozone and CO exposures were not considered by either study, and that the more comprehensive of the two (the ACS study) considered a very limited range of particle metrics constitute severe limitations in the usefulness of the studies.

However, there is an unexpected important implications from these studies (Lipfert, 1995): None of the population subgroups examined appeared to be significantly more sensitive to air pollution than any other (assuming equal probability of

confounding). Since the relative risks were virtually unchanged by excluding subjects with hypertension and diabetes, this finding might also be extended to those with pre-existing chronic diseases. This apparent homogeneity of response has implications regarding the acceptability of population-based studies in which such stratification is not possible. For example, population-based studies such as Lipfert (1993) are unable to incorporate individual risk factors and thus must assume that all members of the population group (city, county, SMSA) are alike in these respects. The results of the prospective studies suggest that this may be a viable assumption.

The two multi-pollutant long-term morbidity studies considered by EPA failed to show any effects on children's respiratory health that could be specifically attributed to fine particles alone. If anything, there seemed to be a slight advantage to the more inclusive PM metrics (such as PM₁₀). Note that, although these particular two studies were limited in scope, they may have offered arguably the best hope of detecting such air pollution morbidity effects, if indeed they exist. They studied children, thus avoiding issues of greatly differing previous community air quality levels and issues of active smoking and occupational exposures. However, like all the other long-term studies, they suffered from lack of personal exposure data and from the potential for regional confounding. The only study to consider bronchitis incidence (as opposed to prevalence) found the lowest odds ratio of the group, implying that prevalence studies may overstate the rate of creation of new cases.

3. Recently Published Epidemiological Studies (not considered by EPA)

3.1. Daily Mortality

3.1.1 Ostro (1996). In this study of mortality in Southern California (Riverside and San Bernadino Counties) daily counts from 1980-86 (mean=41 per day) were regressed against a surrogate measure of PM_{2.5} that was estimated from humidity-adjusted visual range data taken at the local airport (Ontario, CA). Direct measurements of PM_{2.5} were not available. Ostro's regression model controlled for day of week, weekend, year, calendar quarter, month, daily temperature and dew point, but it did not specifically control for hot days (even though the

maximum temperature reached 111°F). Ostro considered several regression model forms, including ordinary least squares, Poisson, deviations from 15-day moving averages, generalized linear modeling, use of trigonometric terms to model seasonality, and stratification by season. Ozone was the only other air pollutant considered; concentrations were measured at four sites in the area; the averaging time used for O₃ was not stated, but since the mean was given as 0.14 ppm, it appears to be the peak hour for each day. The overall mean surrogate PM_{2.5} (based on visibility data) was 32.5 ug/m³, with a maximum of 190 ug/m³. Ostro also reported that an alternative estimation method (based on local measurements) gave a mean of 42.8 ug/m³, with a maximum of 231 ug/m³. Ostro reported that the regression results were insensitive to the use of this alternative PM_{2.5} estimation

method. No other air pollutants were considered, even though data were probably available from the routine monitoring stations that recorded ozone data.

These California counties are downwind of Los Angeles and constitute one of the most polluted areas in the United States, with both O₃ and PM_{2.5} levels more than twice as high as most locations in the Eastern U.S. and in frequent violation of NAAQS. Furthermore, the climate lends itself to frequent outdoor exposures, year round. The daily mortality counts here were high enough so that random noise should not have obscured the air pollution relationships. However, Ostro reported 89% of all deaths in the age 65 and over category, which is suspicious since only about 12% of the population was in that age group. Thus, a relationship between daily mortality and air pollution would be expected to be found here if it in fact existed anywhere.

However, Ostro found that surrogate PM_{2.5} was significantly associated only with total and respiratory mortality, and only in the summer, not with circulatory or deaths of the elderly. This may be first time that the younger age group has been implied to be more important in such a study. Surrogate PM_{2.5} was not significant for the full year for the causes of death considered. Ostro reported that his results were insensitive to the type of regression model used. He only reported ozone results for the summer and for total mortality and did not present a sensitivity analysis for O₃. When regressed as the only pollutant in the model, ozone was significant and had a mean effect of about 2.8% with an upper 95% confidence interval of about 7%. The mean effect of surrogate PM_{2.5} for the same conditions was reported as 3% with an upper confidence interval of 5%. (Since Ostro reported only one digit for these results, the two pollutants cannot be compared precisely, but it appears that O₃ had the larger mean effect in separate regressions.) It is also possible that the failure to specifically consider hot day effects inflated the results for both O₃ and surrogate PM_{2.5}.

Considering O₃ and surrogate PM_{2.5} together in a joint regression, caused both coefficients to decrease and lose significance, which is a typical result with collinear variables. The O₃-surrogate PM_{2.5} correlation in these data was 0.58. In addition, possible differences in exposure error must also be considered when evaluating the results from joint regressions;

when two correlated variables that have equivalent underlying effects are regressed jointly, the one with less measurement error will tend to dominate (Lipfert and Wyzga, 1995a). Ozone is strongly attenuated outdoors by local combustion sources of NO and indoors by reactions on surfaces. In contrast, PM_{2.5} has a long atmospheric residence time and penetrates into buildings where it will remain suspended and may be augmented by indoor sources such as smoking or cooking. In addition, the particular PM_{2.5} data used here were determined from visual range and thus are spatially integrated, which might make them actually perform better in community-wide studies than local measurements would. Thus, the near survival of ozone in the joint regressions could be interpreted as showing a stronger underlying relationship with mortality than surrogate PM_{2.5}. However, in this paper, the

author did not interpret his results in the context of differential measurement error, but suggested that the small difference between mean surrogate PM_{2.5} levels in summer (36 ug/m³) and in winter (29 ug/m³) was responsible for the failure of surrogate PM_{2.5} to reach significance in winter (which implies a threshold). However, the range of this variable was greater in winter (180 ug/m³) than in summer (95 ug/m³), and the range should have been more important in the regression than the mean (unless, of course, a threshold exists).

In this paper, three variables (surrogate PM_{2.5}, O₃, temperature) interacted and were likely intercorrelated to varying degrees; a general correlation matrix was not presented. However, surrogate PM_{2.5} was (just) significant only in summer, when its correlation with temperature was reported to be low. The failure to consider very hot days separately (as has been done by other authors, such as Schwartz and Dockery [1992]) may have overstated the ozone effects. Ostro stated that (because of intercorrelations), "disentangling the effects of temperature versus ozone is problematic". He should have extended this qualification to include the effects of PM_{2.5}.

Ostro noted that the results of his study were consistent with those of Kinney et al. (1991) for Los Angeles County, who also found a stronger effect for O₃ than for PM, but who also noted the difficulty of disentangling PM effects from other pollutants associated with vehicle exhaust (CO and NO₂). It is thus also likely that these problems of collinear pollutants (not considered by Ostro) also existed in Riverside and San Bernadino Counties.

The overall conclusion of this review is that Ostro's study fails to confirm the hypothesis of Schwartz et al. (1996) of an independent effect of fine particles on daily mortality, in spite of the extremely high pollutant levels encountered at the location of study. Ostro's strong results for ozone must be viewed in the context of the frequent violations of the O₃ NAAQS in this location and his inadequate treatment of very hot weather in the regression model.

3.1.2. Anderson et al., 1996. Daily death counts were regressed against air quality for ozone, British smoke, SO₂, and NO₂ in Greater London from 1987 to 1992 by Anderson et al. (1996).

Carbon monoxide was not considered. British smoke measurements generally involve particles of less than 5 μm ; it thus may be considered as a specialized index of fine particles. Anderson et al.'s regression model controlled for season, secular trend, day-of-week, holidays, influenza epidemics, temperature and humidity. They considered deaths from all causes (less accidents, etc.), respiratory diseases, and cardiovascular diseases; the mean daily total death count was 175. Ozone was measured at a single station in central London and was considered as 8-hr daytime averages and daily peak values. NO_2 was measured at two urban sites and was considered as peak hour values and as 24-hr averages. SO_2 and black smoke were measured at four stations that employed the same types of monitoring systems that were used in the 1950s and 60s. the distributions

of black smoke and ozone appeared to be nearly lognormal, with geometric standard deviations (GSDs) of about 1.4 and about 1.6, respectively. The GSDs were somewhat lower for SO₂ and NO₂.

The regression model was based on autoregressive log-linear Poisson regression, following prior work by Schwartz and colleagues (see Schwartz and Dockery, 1992, for example). Trigonometric terms were used for seasonal adjustment; dummy variables controlled for days of the week, holidays, and an influenza epidemic. Temperature effects were represented using a piecewise linear relationship with three segments. Lags up to two days were considered for the air pollutants; each of them was found to maximize the response to at least one of the various pollutants considered. There were 18 combinations of causes of death, pollutants, and averaging times, and each was considered separately for the entire period, the cool season, and the warm season. Thus there were 54 regression coefficients computed, only some of which were actually independent. No non-plausible combinations were considered as "controls." Anderson et al. were somewhat more thorough than many of their predecessors in terms of the range of results presented.

Regression results were reported for the entire period and by season. For the entire period, all-cause mortality was significantly associated with smoke (mean effect = 0.017) and with ozone (8-hr mean effect = 0.014; peak-hr mean effect = 0.017). The mean effects for NO₂ and SO₂ were lower and nonsignificant. No pollutants were significantly associated with cardiovascular deaths, which is perhaps the first time that this finding has been reported. Mean effects on cardiovascular deaths were all less than 0.01. Only ozone was significantly associated with respiratory deaths, with mean effects of 0.035 and 0.042 for 8-hr and peak-hr averages, respectively.

In the warm season, both ozone measures were significantly associated with all three cause of death categories, NO₂ was significant for cardiovascular deaths but negatively significant for respiratory deaths, and SO₂ was significant for all causes. Ozone was significant for respiratory deaths and smoke was significant for all causes in the cool season.

Anderson et al. also presented results for joint pollutant regressions for ozone and smoke for the same combinations of

seasons and causes of death. In these regressions, coefficients for both pollutants generally increased over their separate-regression values, suggesting independent effects. However, this result is also consistent with the negative correlation between the two pollutants. The results for both O₃ and smoke were unaffected by adding SO₂ or NO₂ to the models, but the effects of SO₂ and NO₂ were sensitive to the inclusion of smoke and O₃, respectively. These results are also consistent with their respective bivariate pollutant correlations.

When compared with typical results seen in the United States, both similarities and important differences are seen. Similarities include the facts that the mean air pollution effects in London were similar for all pollutants considered, and

that SO₂ and NO₂ effects were weaker than these for smoke and O₃. In addition, responses for a given pollutant, end point, and season were not reported as a function of lag, so that possible harvesting effects could not be evaluated (like most U.S. studies).

Differences with respect to U.S. studies include the strength of the O₃ effect in comparison with black smoke, although this has only been tested directly in the U.S. in Los Angeles (Kinney and Ozkaynak, 1991). The finding of ozone effects in Central London on a par with those in Los Angeles County, where concentrations are much higher and outdoor exposures may be more common, seems quite unexpected. In addition, a relationship between O₃ and respiratory disease mortality in London winters strains credulity. Black smoke in London may be an index of primary emissions from vehicles, especially diesel engines, whereas summer PM in the U.S. is usually dominated by secondary particles. It is unfortunate that Anderson et al. did not include CO in their study.

The conclusions of Anderson et al. should be viewed in the context of potential differential effects of measurement error. They stated that their ozone readings are undoubtedly biased low because they were obtained in the central city. Ozone values in more outlying areas may also have a more uniform temporal distribution, which could affect their correspondence with mortality peaks. Indoors, both SO₂ and O₃ will be greatly attenuated, but NO₂ and black smoke may be augmented by indoor sources, depending on the fuels used for cooking and heating. Since both CO and smoke are primary emissions from vehicles, the omission of CO from the analysis may have inflated the associations with smoke. In addition, this is one of the few time-series studies that failed to find PM associations with cardiovascular mortality. Finally, Anderson et al. were very thorough, in considering more than one averaging time for O₃ and NO₂ and several causes of death for each pollutant. Their study could have been further improved by considering different age groups and by reporting more details of the lag structures.

The results of Anderson et al. could be used to predict the outcome of a single episode that occurred in London during December 1991, in which peak-hr NO₂ reached 423 ppb and 24-h average smoke reached 228 ug/m³, as reported by Anderson et al.

(1994). The model developed from the winter results of Anderson et al. (1996) would predict about a 25% increase in all-cause mortality, almost all due to smoke; the actual observed increase was only 10%. For cardiovascular mortality, an increase of 14% was observed, but the model would have predicted no net change. For respiratory deaths, the model would have predicted an increase of about 6%; the observed increase was 22%. This unsatisfactory correspondence between predicted and observed changes in daily mortality suggests that models of this type will not be reliable in predicting future improvements in public health that might be gained by eliminating severe air pollution episodes.

3.1.4 The HEI Multi-Pollutant Study of Philadelphia. The Health

Effects Institute (HEI) contracted with Johns Hopkins University to replicate certain of the early acute studies of air pollution and mortality (Samet et al., 1995) and to "explore the sensitivity of findings to analytic approaches and assumptions" using data from Philadelphia (1973-88) as a basis. The first report (Samet et al. 1995) investigated six previously published data sets and generally agreed with the findings of the original investigators, using their data and methods. The report said that "it is reasonable to conclude that, in these six data sets, daily mortality from all causes combined, and from cardiovascular and respiratory causes in particular, increased as levels of particulate air pollution indexes increased". However, they also noted that data from Philadelphia (representing the most deaths under study) suggested that particles were not acting alone.

In the second report (Samet et al., 1996), they found that the results for TSP and/or SO₂ in Philadelphia were relatively insensitive to changes in regression models or to treatments of weather effects. They then added daily data on CO, O₃, and NO₂ and proceeded to a multi-pollutant analysis, starting from scratch with regression model development. They also found that air pollution associations were strongest for those aged 75 and over. In their multi-pollutant model, TSP, SO₂, CO lagged 3 days, and O₃ were significantly positively associated with total mortality, while NO₂ had a significant negative association (a fact that was not mentioned in the abstract). As a single pollutant, NO₂ was not significant (positive), so that the multi-pollutant result probably represented the effects of multicollinearity. Kinney and Ozkaynak (1991) found similar results for Los Angeles when they entered 5 pollutants simultaneously; in this case, the PM variable became negative (but not significant). Similar results were shown by Lipfert and Wyzga (1995a) with simulated data. Nevertheless, the final remarks of Samet et al. (1996) in this report bear repeating (and heeding): "We caution against using the model coefficients directly to estimate the potential consequences of lowering concentrations of the individual pollutants through regulatory measures; the pollutant concentrations are correlated and the estimates of their effects depend on modeling assumptions." It should be perhaps noted that both Samet et al. (1996) and Anderson et al. (1996) have reported significant negative mortality associations with NO₂.

3.1.3 The Birmingham Replication Studies. In 1993, Schwartz published a study of daily mortality and air pollution in Birmingham, AL. This was followed in 1994 by a similar study involving hospital admissions for the elderly. In both cases, Schwartz found that a 3-day average of PM10 provided the maximum implied air pollution effects. Roth and Li presented a replication study of these data at a 1996 CASAC meeting, based on a large number of model variations and lag/pollutant averaging combinations. Although not all of the many combinations that Roth and Li considered were independent, their analysis implied that the claims of Schwartz were not robust and may have represented a statistical artifact. This position was confirmed by a more formal analysis sponsored by the U.S. EPA, reported by Davis et al. (1996). They failed to find a significant

association with PM10 and concluded: "When we use the same variables as included by Schwartz, we obtain results similar to his. But when we use alternative models we obtain different conclusions. In particular, when humidity is included among the meteorological variables (it is excluded in the analysis by Schwartz), we find that the PM10 effect is not statistically significant." Davis et al. went on to say "...model selection is critical in making conclusions about the effect of particulates on mortality, requiring consistent, defensible approaches to assure reliable interpretations".

Moolgavkar et al. (1997) reanalyzed the Birmingham hospital admissions data and found "little evidence of association between air pollution and hospital admissions for respiratory disease". Roth and Li had reported similar findings at the CASAC meeting. The difference between these analyses and that of Schwartz (1994) may well be in the control for differences in admissions by the day of the week. Hospital admissions tend to drop on weekends (this was even true during the major air pollution episode in London in 1952 [Lipfert, 1994]), as do emissions of many primary pollutants. Since these two phenomena are not causally related, failure to control for day-of-week will lead to spurious associations. Since ozone is often higher on weekends (because of reduced titration by NO_x), failure to account for day-of-week can lead to a spurious positive correlation for ozone for a one- or two-day lag (admissions on Mondays correlated with ozone on Saturday or Sunday). This also illustrates the flexibility that the time-series analyst has in seeking the "optimum" model with positive correlations.

3.2 Other Hospitalization and Morbidity Studies

3.2.1 Hospitalization Studies. Several studies have been published recently dealing with daily hospitalization for cardiac and respiratory diagnoses (Morris et al., Schwartz and Morris, Burnett et al.). The U.S. studies are limited to patients aged 65 and over, financed through Medicare. The general conclusion from these studies is that ambient CO is the best predictor for cardiac diagnoses, and O₃ is the best for respiratory diagnoses. Some PM indicators have also been implicated, notably PM10, but the relative performance of these pollutants in joint regressions must be viewed in the perspective of exposure error. O₃ is strongly attenuated indoors (reference needed), and its outdoor

distribution in urban areas is affected by chemical reactions (titration) with local sources of NO_x. Since CO in urban areas has only two major sources, wood burning and vehicles, its outdoor distributions tend to be much more variable than those for the smaller particles fractions of PM (reference needed). Both CO and O₃ tend to be highly correlated with PM (in winter and summer, respectively). Thus, the fact that both CO and O₃ perform better than PM in joint regressions, in spite of their typically higher exposure errors, constitute evidence that CO and O₃ are more likely to evoke health responses than PM, at current ambient levels. This conclusion follows directly from the simulation studies of Lipfert and Wyzga (1995a).

3.2.2 Emergency Room Visits for Respiratory Illnesses in

Montreal. Delfino et al. (1997) tracked ER visits for respiratory illness for 25 hospitals in Montreal for the summers of 1992 and 1993. Only the visits that required medical treatment and monitoring on a stretcher were included in the analysis. A control group was defined consisting of non-ambulatory patients with gastrointestinal or psychiatric complaints. Seasonal and day-of-week effects were controlled. The air pollutants considered included ozone, PM₁₀, PM_{2.5}, SO₄⁼, and strong acidity (H⁺); all were measured daily, ozone at 7 stations. The results showed the largest mean effect for ozone (21%), with PM₁₀, PM_{2.5}, SO₄⁼, and acidity (not significant) following in order, based on single-pollutant regressions. The mean effect of SO₄⁼ was about 6%, and the authors showed that this result was reasonably consistent with several previous studies. Only PM₁₀ was significantly associated with the control admissions (at lag 0), but this raises doubts about the possibility of spurious results for the other endpoints as well.

3.2.3 A Swiss Cross-Sectional Study of Lung Function. Ackermann-Lieblich (1997) examined about 10,000 Swiss adults living in 8 different locations, including two at higher altitudes. Lung function and allergy tests were conducted and air quality data were obtained from local routine measurements (SO₂, NO₂, TSP, PM₁₀, and O₃). Analyses were done separately for by smoking status and presence of respiratory symptoms. The authors reported that "significant and consistent effects" on lung function were found for SO₂, NO₂, and PM₁₀, for all subgroups. Results were not reported for TSP, but since the correlation between TSP and PM₁₀ was 0.95, one would expect the lung function associations to be similar. They declined to try to distinguish the effects of the various pollutants. However, some of the lung function differences due to air pollution lost significance when altitude was included as a variable. The magnitude of the effect attributed to PM₁₀ (3.4% decrease in FVC for 10 ug/m³ in PM₁₀) seems quite large in comparison to other studies.

4. Studies That EPA Should Have Considered in Their Proposal But Did Not.

EPA selected particular studies from the more inclusive Criteria Document for use in the Staff Paper and Decision Document. Some of the equally valid studies that present contrasting views have

been discussed above. Others are discussed in the following section.

4.1 Overlooked Short-Term Mortality Studies

4.1.1 Review and Reconciliation of Results. Lipfert and Wyzga (1995b) reviewed 27 time-series mortality studies that imply that about 1-8% of daily deaths may be associated with daily variations in air quality, but the design of these studies does not allow estimates of the degree of prematurity of death. There are also uncertainties about the "responsible" pollutants, because of the fundamental collinearity among time-varying pollutants. One of the methodologies that focuses on longer-term effects is that of the cross-sectional analysis, which deals with

spatial relationships. Lipfert (1995) considered and compared the two types of cross-sectional studies: prospective cohort studies, which deal with survival of individuals, and population-based "ecological" studies that analyze differences among communities. Both types relied on community-based air monitoring data and thus are "ecological" in that sense.

The published cross-sectional studies of long-term associations between air pollution and mortality identified several obstacles to precise determination of relationships. These include uncertainties as to the "correct" regression models, personal lifestyle risk factors for which appropriate data may not be available, changes in heart disease mortality (resulting in part from changes in the prevalence of these risk factors) that differ by region, collinearity among pollutants, and unavailability of appropriate data on certain pollutants such as CO. In addition, differences in historical air pollution levels that might be associated with chronic responses have usually been ignored. However, one of the outcomes from the prospective studies has been the failure to identify any particularly susceptible population subgroup that might affect the plausibility of the ecological assumption (i.e., that the behavior of individuals may be inferred from data on groups). As a result, there appear to be no reasons why the results of prospective cohort studies should not be compared to those from ecological studies.

Two such types of comparisons appear to be relevant at this time. First, all of the published cross-sectional studies that considered air pollution concurrent with mortality found significant associations; however, they differed by pollutant and by magnitude of response, and some of these differences appeared to be due to differences in the degrees to which potential confounding variables were controlled by the various models. Next, the relative risks derived by the prospective cohort studies (Dockery et al., 1993; Pope et al., 1995) appear to substantially exceed those of the daily mortality studies, which might be interpreted as either showing the existence of effects due to chronic exposures to air pollution or the presence of regional confounding. However, a portion of the "excess" relative risks reported by the prospective studies may be a result of using concurrent rather than historic or cumulative air pollution exposure levels, which are substantially higher for the most polluted locations. Substituting the higher exposure levels

would bring the estimates from the prospective studies more into line with the ecological cross-sectional studies and with the daily mortality studies.

Although there are good reasons for comparing long- and short-term exposure studies, the power of such comparisons is limited by the fundamental uncertainties of both types of studies. There are lingering questions about the appropriate lag structures, seasonal adjustments, and weather variables in the daily time series studies (Wyzga and Lipfert, 1995) as well as with the duration of peak air pollution exposures. As discussed above, the long-term studies are sensitive to the regression models used, the degree of confounder control, and the lag structure assumed for long-term exposures (annual averages). In

addition, there is no a priori reason to assume that the same pollutants would be responsible for both acute and chronic responses. The combined effects of all of these uncertainties appear to be of the same order of magnitude as the net differences between types of studies, making the establishment of chronic effects on this basis problematic.

It would thus appear that additional methods must be developed if reliable estimates of the long-term effects of air pollution are to be derived. The first requirement should be the establishment of specific disease-pollutant hypotheses to be tested, preferably on the basis of clinical theory and laboratory experiments. The relevant epidemiological data should probably then be obtained from the records of long-term pollution abatement histories. Such a space-time analysis might succeed if enough contrasting situations can be identified.

The effect on the regression coefficients of using lagged air pollution data has already been discussed. This final comment is addressed to the suggestion by Pope et al. (1995) that the ecological relationships between air pollution and age-race-sex-adjusted mortality data may be used to justify relative risks of the order of 1.25. Figure 3 is a plot of SMSA mortality against SO₄²⁻, ca. 1980, for various adjustment procedures. The plot on the left resembles Figure 1 of Pope et al. on which this suggestion was based, although the exact correspondence between the locations considered could not be established. The plot in the center includes mortality adjustments for college education and for smoking and being overweight (state-level data for the latter two parameters). This corresponds roughly to the actual Six City analysis, and the slope is reduced somewhat. The plot on the right extends the mortality adjustments to a larger range of variables that have been shown exogenously to influence mortality; this slope is essentially zero, consistent with the regression results. Figure 4 extends this analysis to PM_{2.5}, for a smaller number of locations. Here the slope reduction due to introducing additional nonpollution variables is less drastic but still substantial. These plots are intended to demonstrate that relationships between longevity and air pollution based on spatial gradients cannot be established with confidence without considering a full range of possible confounders. This precept applies to prospective cohort and population-based (ecological)

studies alike.

4.1.2 Air Pollution and Mortality in Amsterdam. Verhoef et al. (1996) studied the associations of all-cause mortality (mean daily count=19) in Amsterdam with several indices of urban air pollution from 1986-92. They considered measurements at 3 population oriented sites for smoke, PM10 measured every 3-4 days, SO₂, CO and O₃. Daily averages were used for all pollutants except ozone, for which peak hourly averages were used. A prediction algorithm was used to fill in the missing PM10 data. Weather data included temperature and relative humidity from the airport. Data on flu epidemics were also considered; day-of-week and temporal trends were also controlled. Lags up to 2 days were considered. All average air concentrations were low: smoke, 12

ug/m³; estimated PM₁₀, 38 ug/m³; SO₂, 5 ppb; CO, 0.85 ppm; O₃, 22 ppb. Regression results were presented for each pollutant and lag; it is not clear if the results for each lag are independent. The results were presented as relative risks for 100 ug/m³ of each pollutant, which is very misleading because of the great disparity in mean values. Statistical significance was only obtained for smoke and ozone, but the mean effects were quite similar for all pollutants except and SO₂, which was lower. When summed over all lags, the mean effects were: smoke, 0.044; estimated PM₁₀, 0.030; SO₂, 0.014; CO, 0.03*; O₃, 0.029. When combined with gaseous pollutants in 2-pollutant regressions, the PM measures retained significance at the expense of the gases. The authors concluded that black smoke was the best predictor of daily mortality in their study, in part because of particle size but also because of its chemical constituents. It should be noted that these constituents have little in common with the soluble sulfate compounds currently being emphasized in the U.S. Also, possible effects of differential measurement error should be considered here, given that a large percentage of the PM₁₀ values were estimated.

4.2 Long-Term Mortality Studies

4.2.1 Prospective Cohort Study of California Seventh-Day Adventists. Abbey et al. (1991) described a prospective study of about 6000 white, non-Hispanic, nonsmoking, long-term California residents who were followed for 6-10 years, beginning in 1976. There were 845 deaths from non-external causes. The study was designed to test the use of cumulative population-oriented air quality data as an explanatory factor for disease incidence and chronic effects. Ambient air quality data (TSP, O₃) dating back to 1966 were used, and the study was restricted to those who lived within 5 miles of their current residence for at least 10 years. The follow-up analysis (Abbey et al., 1995) considered exposures to SO₄²⁻, PM₁₀ (estimated from site-specific regressions on TSP), PM_{2.5} (estimated from visibility), and visibility per se (atmospheric extinction coefficient). All of the air quality monitors in the state were used to create individual exposure profiles (duration of exposure to specific minimum concentration levels) for each participant, by interpolating to their zip code centroids based on the three nearest monitoring stations. The Cox proportional hazards model was used, considering age, sex, past smoking, education, and

presence of definite symptoms of airway obstructive disease (asthma, chronic bronchitis, or emphysema) in 1977 as individual risk factors, together with various exposure indices for air pollution, considered separately. Data on occupational exposures and history of high blood pressure were available but were not used in the mortality model. No data were available on climate, body mass, income, migration, physical activity levels or diet and separate results were not reported for mortality by gender.

Neither heart attacks or nonexternal mortality was associated with any pollutant in these studies. The authors felt that possible errors in their estimated exposures to air pollution may have contributed to the lack of significant findings. The finding of no association between long-term cumulative exposure

to air pollution and mortality may be interpreted as showing the absence of chronic responses after 10 years but not necessarily the absence of (integrated) acute responses, since coincident exposures were not considered. It is also possible that the latency period for chronic effects may exceed 10 years and that additional follow-up might still reveal chronic effects. The imputed magnitudes of the other risk factors considered were not given.

4.2.2 Studies of 1980 City Mortality. The analysis by Lipfert et al. (1988) comprised a statistical analysis of 1980 spatial patterns among U.S. central cities for total mortality (all causes), evaluating demographic, socioeconomic, and air pollution factors as predictors. The advantages of studying central cities instead of SMSAs include potentially better measures of air pollution exposure because of the smaller land areas and higher population densities, and larger numbers of observations that allow analysis of subsets. Lipfert et al. found sulfate* and iron particles to be significant predictors of all-cause mortality in about 180 cities; TSP was considered only cursorily. If the elasticities for SO₄²⁻ were corrected to account for the filter artifacts, they would be reduced by about 50% in this analysis (Lipfert, 1994a). The data on PM₁₅ and PM_{2.5} were only available for 68 cities; neither pollutant was statistically significant in this data set but their elasticities were in the same range found for other pollutants (0.013-0.05). Lipfert et al. also tested lagged pollution data as a means of attempting to distinguish acute from chronic responses; using ca. 1970 TSP and SO₄²⁻ data to predict 1980 city mortality was slightly less effective than using ca. 1980 data for these pollutants as predictors. This result does not suggest the presence of chronic effects.

* CO results are uncertain because only 1 digit was presented in the tabulated results.

4.2.3 Studies of 1980 SMSA Mortality. Ozkaynak and Thurston (1987) analyzed 1980 total mortality in 98 SMSAs against EPA AIRS data on TSP and SO₄²⁻ and data on PM₁₅ and PM_{2.5} from the EPA inhalable particle (IP) monitoring network for 38 of these locations. The sulfate measurements that Ozkaynak and Thurston used were probably affected by artifacts from the high-volume sampler filters (Lipfert, 1994a). The sulfate data from the IP

network were not used by Ozkaynak and Thurston.

Ozkaynak and Thurston ranked the importance of the various pollutants mainly by relative statistical significance in separate regressions. They concluded that the results were "suggestive" of an effect of particles on mortality decreasing with particle size, although in the basic model only SO₄²⁻ was significant. In some of the other models, PM_{2.5} was also significant and PM₁₅ was nearly so. However, if the effects are judged by elasticities rather than significance levels, SO₄²⁻, PM_{2.5}, and PM₁₅ would be judged as equivalent, with TSP ranking somewhat lower. The indicated effect of SO₄²⁻ would be reduced from an elasticity of 0.086 to about 0.05 by accounting for filter artifacts. Ozkaynak and Thurston also used source

apportionment techniques to estimate that particles from coal combustion and from the metals industry appeared to be the most important.

The coefficients and significance levels obtained for TSP by Ozkaynak and Thurston may be the result of the data they used, which were based on a single monitoring station in each SMSA and thus are unlikely to be fully representative of average population exposures to TSP in areas as large as SMSAs. As a result, alternative interpretations of these findings are certainly possible.

In addition, because smoking, diet, and other socioeconomic or lifestyle variables were not considered in the regression model, the pollution coefficients may have been biased. Finally, Ozkaynak and Thurston did not specifically address the question of acute vs. chronic responses by exploring lagged pollution variables.

Lipfert (1993) used data from up to 149 metropolitan areas (mostly SMSAs) in a study of the relationships between community air pollution and "excess" mortality due to various causes for the year 1980. Several socioeconomic models, including the model proposed by Ozkaynak and Thurston (1987), were used in cross-sectional multiple regression analyses to account for non-pollution effects. Cause-of-death categories analyzed include all causes, nonexternal causes (ICD-9 0-800), major cardiovascular diseases (ICD-9 390-448), and chronic obstructive pulmonary diseases (COPD) (ICD-9 490-96). The spatial patterns for the first three groupings were quite similar but differed markedly from the patterns of COPD mortality, which tend to be higher in the Western U.S. Regressions were performed for these cause-of-death groupings as annual mortality rates ("linear" models) and as their logarithms ("log-linear" models).

Two different sources of measured air quality data were utilized. Data from the EPA AIRS database included TSP, SO₄, Mn, and ozone (obtained from a long-term average isopleth map). Data from the inhalable particulate network included PM₁₅, PM_{2.5} and SO₄ from the teflon IP filters for 63 locations. All particulate data were averaged across all the monitoring stations available for each SMSA; the TSP data were restricted to the year 1980 and were based on an average of about 10 sites per SMSA.

This represents a substantial improvement over previous air quality representations.

* The sulfate data (along with data on ambient SO₂ and NO_x) that yielded the best results were obtained from a long-range transport model (Shannon, 1981), instead of from local measurements. This lack of measurement error may have contributed to the success of these three variables, which were virtually indistinguishable from one another. However, Shannon has since reported in a personal communication the discovery of a systematic error in the meteorological data used as input to the model. This error biases some of the results high. For this reason, the findings of Lipfert et al. (1988) with respect to

SO₄-, SO₂, and NO_x may be unreliable.

The associations between mortality and air pollution were found to be dependent on the socioeconomic factors included in the models, the specific locations included in the data set, and the type of statistical model used, as was the case with a previous analysis of 1970 data (Lipfert, 1978). An important variable that interacts with indices of industrial air pollution is the rate of population change, which is negatively correlated with mortality and with sulfur air pollution.

Stepwise regressions were run for each mortality variable, and a "parsimonious" model was developed that had statistically significant coefficients for the non-pollution variables. Most of these coefficients also agreed with exogenous estimates of the "correct" magnitudes of the risks. Using these models, statistically significant associations were found between TSP and mortality due to non-external causes with log-linear models, but not with linear models. Sulfates, manganese, inhalable particles (PM₁₅), and fine particles (PM_{2.5}) were not significantly ($p > 0.05$) associated with mortality with any of the parsimonious models, although PM_{2.5} and Mn were close with linear models ($p = 0.07$) and statistical significance may have been affected by the use of smaller data sets for these species. PM_{2.5} was a significant predictor of heart disease mortality only when the regression model was restricted to the variables used by Ozkaynak and Thurston (1987). Lipfert showed that PM_{2.5} was the "strongest" particulate variable with linear models, but that TSP performed better in log-linear models. In that sense, the findings for PM were not robust. Scatter plots and quintile analyses suggested that a TSP threshold might be present for COPD mortality, at around 65 $\mu\text{g}/\text{m}^3$ (annual average).

Lipfert's results provided some support for previous findings of associations between TSP and premature mortality but not for fine particles. They also (indirectly) supported the hypothesis that improving the accuracy of pollutant exposure data tends to increase statistical significance. Similarly, the lack of significance for SO₄ may partly relate to the flawed measurement methods used at the time. The ambiguity between linear and log-linear models probably reflects the effects of influential observations. However, the study failed to consider other socioeconomic and lifestyle variables (such as diet and exercise

habits) that could have been confounded with air pollution.

4.2 Overlooked Morbidity Studies

4.2.1 The NHANES II Data on Lung Function. Schwartz's (1989) observational study is in many ways a paradigm for deriving estimates of the effects of air pollution on lung function. The lung function data were obtained as part of the second National Health and Nutrition examination Survey (NHANES), which was intended to be a random, national sample of the civilian, noninstitutionalized population of the United States. The NHANES II spirometry data were limited to children and young adults (ages 6-24). Care was taken to ensure standardization in the collection and processing of the data, which were acquired from 1976 to 1980. Schwartz based his analysis on the best trial for each subject and deleted data that did not meet spirometry standards. The lung function data (FVC, FEV1, and PEF) were fitted to regression models based on race, age, sex, number of cigarettes smoked per day, and several physiological measures. The residuals from these regressions were then regressed against air pollution, one pollutant at a time, as a 2-stage procedure. The possibility of neighborhood clustering* was accounted for by using a nested random effects model similar to that of Ware et al. (1986).

Air pollution data were obtained from the EPA's air quality data base (now called AIRS, Aerometric Information Retrieval System), using only those monitors that were located within 10 miles of the population centroid of the census tract of the subject's residence. This condition resulted in different numbers of acceptable cases for each pollutant considered, as shown in Table 11-2. Temporal averaging consisted of the 365 days preceding the examination. The ozone regressions were based on the annual average of daylight hours (8-hour averages). This statistic is about 50 to 100% greater than the annual average of all hours.

Schwartz found statistically significant negative associations between each of the three lung function metrics and NO₂, O₃, and TSP, but not for SO₂. (One of the problems with national data on SO₂ may be the variability among measurement methods in use, especially during this period.) The strongest associations were for ozone and FVC. Robustness of the results was established by re-estimating the relationships with non-linear pollution transforms, including thresholds; by excluding subjects with chronic respiratory conditions, smokers, subjects not residing in

the state of their birth; by including factors for family socioeconomic status, urbanization, region of the country; by using 2-year means for pollution data; and by extending the allowable monitor radius to 20 miles. Excluding subjects with chronic respiratory conditions reduced all of the pollution coefficients; excluding smokers increased them. Increasing the radius for the pollution monitors reduced only the TSP coefficients. All of these findings are consistent with a causal hypothesis. Schwartz also noted that linear dose-response models for the air pollutants appeared to be superior to logarithmic transforms.

The nonlinear dose-response relationships suggested a threshold for ozone of about 0.04 ppm for the daylight average (0.02-0.025

ppm annual average); this is a relatively high value (90th percentile in Schwartz's data), which might correspond to average daily 1-hour peaks around the current NAAQS of 0.12 ppm (235 ug/m³). For TSP, a threshold was suggested around 90 ug/m³, which exceeds the current standard of 75 ug/m³. The NO₂ relationship was more nearly linear, with an increase in slope at about 0.04 ppm (75 ug/m³), which is below the current federal standard. These results suggest that most of the "signal" in Schwartz's data set was derived from the locations above the 90th percentile in air pollution, for all species. The lack of significant findings for SO₂ is also noteworthy, since the 90th percentile for SO₂ was 0.019 ppm (50 ug/m³) and there are relatively few urban locations in the United States which now exceed this value (see Table 2-8, for example). Association with SO₄= aerosol was not evaluated, but regional dummy variables were investigated that may provide some insight, given the regional distribution of this pollutant. No dummy variable was significant for FEV1 or

* Cohen (1980) and others have explored the extent to which a genetic deficiency contributes to impaired pulmonary function. To the extent that close relatives tend to live in the same neighborhood, community, or region, confounding could thus occur between local environmental effects and the effects of shared genetic defects. This phenomenon could also help explain the apparent geographic clustering in the 24 Cities Study discussed above.

FVC, but peak flow was about 5% lower in the Northeast. Schwartz also reported decrements in lung function associated with central city residence, apart from the specific air pollution relationships studied. Such a relationship would imply a primary pollutant (if any), rather than a secondary species like SO₄= or PM_{2.5}.

4.2.2 Acid Aerosols and Respiratory Symptoms. Schwartz et al. (1989) followed a cohort of about 300 children selected from the Six Cities Study, whose families kept diaries for about a year, beginning in the fall of 1984. Respiratory symptoms were recorded and each family was called biweekly and mailed their diary to the investigators monthly. Daily measurements of SO₂, NO₂, O₃, PM₁₀, PM_{2.5}, atmospheric visibility, SO₄=, sulfuric acid, and aerosol strong acidity were made. 48-hr air quality averages were used in the regressions. For logistical reasons,

the study extended over four school years (1984-88). This paper emphasized "lower respiratory illness", defined as presence of at least two symptoms among cough, chest pain, phlegm, or wheeze. An event was defined as an occurrence of new symptoms. The authors reported significant results for all pollutants except the acids. The ranking of the significant pollutants by mean effect was, in decreasing order: PM10 and O3 (tied for first), PM2.5, SO4=, visibility, and SO2 (distant last). A dose-response plot for O3 showed approximate linearity, with an approximate doubling of the incidence of lower respiratory illness (from 0.07% to 0.14%) as ozone increased from about 20 ppb to about 60 ppb. Temperature was also found to be an important predictor of respiratory symptoms.

4.2.3 A Study of Respiratory Outpatient Visits in Alaska. Gordian et al. (1996)* analyzed outpatient visits to both physicians' offices and hospital emergency rooms from May 1992 to March 1994, in Anchorage, Alaska. Air quality measurements came mainly from a central site for PM₁₀. CO was measured during winter months at five sites. The main sources of suspended particulates were reported to be crustal materials and volcanic ash. The data were adjusted for day-of-week effects and filtered to control for seasonality. The regression results showed a mean effect of PM₁₀ of 13-27% on asthma visits and 4-13% for upper respiratory visits. The authors concluded that "the coarse fraction of PM₁₀ may affect the health of working people".

However, CO was found to be associated with visits for bronchitis and for upper respiratory conditions, which raises questions about biological plausibility for this and other similar studies. The direct effects of CO are thought to be limited to the oxygen-carrying capacity of the blood, so that associations with respiratory symptoms must reflect some sort of surrogate effects. All automobile-related pollutants (CO, aldehydes, soot, NO_x, VOCs) tend to be highly correlated in time, so that any of these species could be the actual causal agent for respiratory effects, as also noted in the acute mortality study of Ozkaynak and Kinney (1991) in Los Angeles. By the same token, findings of associations between any one of these species and cardiac complaints should be interpreted as applying to any member of the group. This reinforces the ambiguity between PM and CO that has been noted throughout this review.

Finally, it should be noted that there is no evidence in the Alaska study implicating sulfates, aerosol acidity, or any other soluble fine particle with respiratory morbidity.

5. Important Technical Issues Affecting the Scientific Validity of EPA's Proposal

5.1 Issues in Time-Series Studies

5.1.1. Seasonal Adjustment and Lag Selection. In addition to daily variations in the timing of health-related events, there are also large seasonal variations that must be removed or "controlled", as well as differences that occur by day of the week and on holidays.*** There are many ways of trying to do

this, some better than others. After such treatment, the air quality and health data can be plotted as series of peaks and valleys oscillating about a horizontal line.

** A preliminary version of this paper was made available to EPA during preparation of the CD.

The extent to which the health and pollution peaks match up defines their implied statistical relationship. A certain amount of matching is to be expected just due to chance, and if the lag between exposure and response is free to vary, it becomes much easier to find such a match (Pickles, 1981). Averaging the pollution data over more than one day also improves the probability of matching.

During the major air pollution episodes of past decades (which form the paradigm for time-series analyses), health responses appeared to track pollution excursions throughout the periods of poor air quality. By selecting an appropriate lag between pollution excursion and response, a near-linear dose-response function could be hypothesized that included both increases and decreases over several days. This kind of matching gives rise to correlations around 0.5 or more. In contrast, the vast majority of current situations involve peaks for single days only which are thus shaped as triangles; the shapes or temporal profiles of the excursions thus play little or no role in the analysis. This means that the correlations between excursions and health responses depend on the numbers of triangles that appear to coincide with appropriate lags; the shapes or durations of episodes are no longer a factor. These correlations are much lower (ca. 0.1) and depend on a large number of repeating cycles to achieve significance. Examination of episodes of high pollution and of high mortality in contemporary atmospheres shows little or no correspondence; this can only be explained by either substantial measurement error (people are not exposed to the highest concentrations) or by (undefined) statistical artifacts in the analysis.

5.1.2. Pollutant and Weather Collinearity. All air pollutants that are generated locally fluctuate similarly to changes in weather, such as with wind speed. Pollutants that emanate from a common source may share even more characteristics; examples are the particles and gases released as combustion products. However, monitoring networks are not the same for all pollutants and these uncertainties will tend to mask the degree of collinearity actually present among pollutants. For example, in Philadelphia, we find that SO₂, NO₂ and CO track PM nearly perfectly (Figures 5-10, based on data from Wyzga and Lipfert, 1995) after averaging to reduce measurement variability. Daily change in temperature also tracks PM in this way (Figure 11). Regional pollutants, such as ozone, sulfates, and fine particles, only partially track the other pollutants because of their dependence on photochemistry; however, these pollutants also tend to track weather phenomena, but differently than primary pollutants. The ability to distinguish these collinear pollutants from one another depends on their relative measurement errors.

EPA and others seem to be relying upon spatial contrasts among

time-series results to address the pollutant collinearity issue. EPA cites the finding of similar associations between PM and daily mortality in locations with both high and low levels of sulfur oxides as evidence that PM is the agent and not SO_x, for example. However, there are important gaps in this line of reasoning. The most common co-pollutant that has been considered with PM is SO₂; these two species tend to be highly correlated and, since SO₂ has a greater exposure uncertainty (because it is highly attenuated indoors), PM is usually the "winner" in a joint regression. Carbon monoxide is a much more likely confounder of PM and has only been considered in a few instances. Kinney and Ozkaynak (1991) found that they could not distinguish among PM,

*** Such effects can give rise to spurious correlations between